The Bell System Technical Journal

January, 1934

Stabilized Feedback Amplifiers*

By H. S. BLACK

This paper describes and explains the theory of the feedback principle and then demonstrates how stability of amplification and reduction of modulation products, as well as certain other advantages, follow when stabilized feedback is applied to an amplifier. The underlying principle of design by means of which singing is avoided is next set forth. The paper concludes with some examples of results obtained on amplifiers which have been built employing this new principle.

The carrier-in-cable system dealt with in a companion paper [1] involves many amplifiers in tandem with many telephone channels passing through each amplifier and constitutes, therefore, an ideal field for application of this feedback principle. A field trial of this system was made at Morristown, New Jersey, in which seventy of these amplifiers were operated in tandem. The results of this trial were highly satisfactory and demonstrated conclusively the correctness of the theory and the practicability of its commercial application.

Introduction

DUE TO advances in vacuum tube development and amplifier technique, it is now possible to secure any desired amplification of the electrical waves used in the communication field. When many amplifiers are worked in tandem, however, it becomes difficult to keep the overall circuit efficiency constant, variations in battery potentials and currents, small when considered individually, adding up to produce serious transmission changes for the overall circuit. Furthermore, although it has remarkably linear properties, when the modern vacuum tube amplifier is used to handle a number of carrier telephone channels, extraneous frequencies are generated which cause interference between the channels. To keep this interference within proper bounds involves serious sacrifice of effective amplifier capacity or the use of a push-pull arrangement which, while giving some increase in capacity, adds to maintenance difficulty.

However, by building an amplifier whose gain is deliberately made, say 40 decibels higher than necessary (10,000 fold excess on energy basis), and then feeding the output back on the input in such a way

* Presented at Winter Convention of A. I. E. E., New York City, Jan. 23–26, 1934. Published in *Electrical Engineering*, January, 1934.

[1] "Carrier in Cable" by A. B. Clark and B. W. Kendall, presented at the A. I. E. E. Summer Convention, Chicago, Ill., June, 1933; published in *Electrical Engineering*, July, 1933, and in *Bell Sys. Tech. Jour.*, July, 1933.

as to throw away the excess gain, it has been found possible to effect extraordinary improvement in constancy of amplification and freedom from non-linearity. By employing this feedback principle, amplifiers have been built and used whose gain varied less than 0.01 db with a change in plate voltage from 240 to 260 volts and whose modulation products were 75 db below the signal output at full load. For an amplifier of conventional design and comparable size this change in plate voltage would have produced about 0.7 db variation while the modulation products would have been only 35 db down; in other words, 40 db reduction in modulation products was effected. (On an energy basis the reduction was 10,000 fold.)

Stabilized feedback possesses other advantages including reduced delay and delay distortion, reduced noise disturbance from the power supply circuits and various other features best appreciated by practical designers of amplifiers.

It is far from a simple proposition to employ feedback in this way because of the very special control required of phase shifts in the amplifier and feedback circuits, not only throughout the useful frequency band but also for a wide range of frequencies above and below this band. Unless these relations are maintained, singing will occur, usually at frequencies outside the useful range. Once having achieved a design, however, in which proper phase relations are secured, experience has demonstrated that the performance obtained is perfectly reliable.

Circuit Arrangement

In the amplifier of Fig. 1, a portion of the output is returned to the input to produce feedback action. The upper branch, called the μ-circuit, is represented as containing active elements such as an amplifier while the lower branch, called the β-circuit, is shown as a passive network. The way a voltage is modified after once traversing each circuit is denoted μ and β respectively and the product, $\mu\beta$, represents how a voltage is modified after making a single journey around amplifier and feedback circuits. Both μ and β are complex quantities, functions of frequency, and in the generalized concept either or both may be greater or less in absolute value than unity.[2]

Figure 2 shows an arrangement convenient for some purposes where, by using balanced bridges in input and output circuits, interaction between input and output is avoided and feedback action and amplifier impedances are made independent of the properties of circuits connected to the amplifier.

[2] μ is not used in the sense that it is sometimes used, namely, to denote the amplification constant of a particular tube, but as the complex ratio of the output to the input voltage of the amplifier circuit.

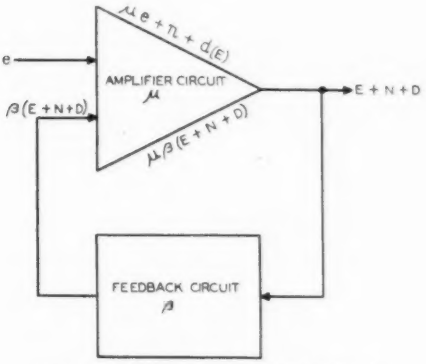


Fig. 1—Amplifier system with feedback.

e—Signal input voltage.
μ—Propagation of amplifier circuit.
μe—Signal output voltage without feedback.
n—Noise output voltage without feedback.
$d(E)$—Distortion output voltage without feedback.
β—Propagation of feedback circuit.
E—Signal output voltage with feedback.
N—Noise output voltage with feedback.
D—Distortion output voltage with feedback.

The output voltage with feedback is $E + N + D$ and is the sum of $\mu e + n + d(E)$, the value without feedback plus $\mu\beta[E + N + D]$ due to feedback.

$$E + N + D = \mu e + n + d(E) + \mu\beta[E + N + D]$$

$$[E + N + D](1 - \mu\beta) = \mu e + n + d(E)$$

$$E + N + D = \frac{\mu e}{1 - \mu\beta} + \frac{n}{1 - \mu\beta} + \frac{d(E)}{1 - \mu\beta}$$

If $|\mu\beta| \gg 1$, $E \doteq -\frac{e}{\beta}$. Under this condition the amplification is independent of μ but does depend upon β. Consequently the over-all characteristic will be controlled by the feedback circuit which may include equalizers or other corrective networks.

General Equation

In Fig. 1, β is zero without feedback and a signal voltage, e_0, applied to the input of the μ-circuit produces an output voltage. This is made up of what is wanted, the amplified signal, E_0, and components that are not wanted, namely, noise and distortion designated N_0 and D_0 and assumed to be generated within the amplifier. It is further assumed that the noise is independent of the signal and the distortion generator or modulation a function *only of the signal output.* Using the notation of Fig. 1, the output without feedback may be written as:

$$E_0 + N_0 + D_0 = \mu e_0 + n + d(E_0), \tag{1}$$

where zero subscripts refer to conditions without feedback.

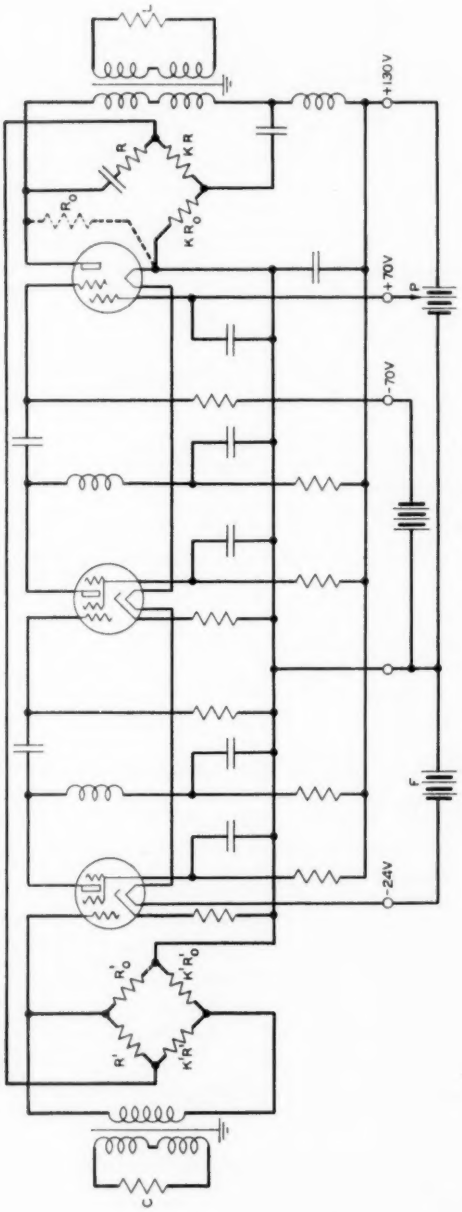


Fig. 2—Circuit of a negative feedback amplifier.

With feedback, β is not zero and the input to the μ-circuit becomes $e_0 + \beta(E + N + D)$. The output is $E + N + D$ and is equal to $\mu[e_0 + \beta(E + N + D)] + n + d(E)$ or:

$$E + N + D = \frac{\mu e_0}{1 - \mu\beta} + \frac{n}{1 - \mu\beta} + \frac{d(E)}{1 - \mu\beta}. \tag{2}$$

In the output, signal, noise and modulation are divided by $(1 - \mu\beta)$, and assuming $|1 - \mu\beta| > 1$, all are reduced.

Change in Gain Due to Feedback

From equation (2), the amplification with feedback equals the amplification without feedback divided by $(1 - \mu\beta)$. The effect of adding feedback, therefore, usually is to change the gain of the amplifier and this change will be expressed as:

$$G_{CF} = 20 \log_{10} \left| \frac{1}{1 - \mu\beta} \right|, \tag{3}$$

where G_{CF} is *db change in gain due to feedback*. $1/(1 - \mu\beta)$ will be used as a quantitative measure of the effect of feedback and the feedback referred to as *positive feedback* or *negative feedback* according as the absolute value of $1/(1 - \mu\beta)$ is greater or less than unity. Positive feedback increases the gain of the amplifier; negative feedback reduces it. The term *feedback* is not limited merely to those cases where the absolute value of $1/(1 - \mu\beta)$ is other than unity.

From $\mu\beta = |\mu\beta| \angle \Phi$ and (3), it may be shown that:

$$10^{-G_{CF}/10} = 1 - 2|\mu\beta| \cos \Phi + |\mu\beta|^2, \tag{4}$$

which is the equation for a family of concentric circles of radii $10^{-G_{CF}/10}$ about the point 1, 0. Figure 3 is a polar diagram of the vector field of $\mu\beta = |\mu\beta| \angle \Phi$. Using rectangular instead of polar coordinates, Fig. 4 corresponds to Fig. 3 and may be regarded as a diagram of the field of $\mu\beta$ where the parameter is db change in gain due to feedback. From these diagrams all of the essential properties of feedback action can be obtained such as change in amplification, effect on linearity, change in stability due to variations in various parts of the system, reduction of noise, etc. Certain significant boundaries have been designated similarly on both figures.

For example, boundary A is the locus of zero change in gain due to feedback. Along this parametric contour line where the absolute magnitude of amplification is not changed by feedback action, values of $|\mu\beta|$ range from zero to 2 and the phase shift, Φ, around the amplifier

and feedback circuits equals $\cos^{-1} |\mu\beta|/2$ and, therefore, lies between $-90°$ and $+90°$. For all conditions inside or above this boundary, the gain with feedback is increased; outside or below, the gain is *decreased*.

Stability

From equation (2), $\mu e_0/(1 - \mu\beta)$ is the amplified signal with feedback and, therefore, $\mu/(1 - \mu\beta)$ is an index of the amplification. It is of course a complex ratio. It will be designated A_F and referred to as the amplification with feedback.

To consider the effect of feedback upon stability of amplification, the stability will be viewed as the ratio of a change, δA_F, to A_F where δA_F is due to a change in either μ or β and the effects may be derived by assuming the variations are small.

$$A_F = \frac{\mu}{1 - \mu\beta}, \tag{5}$$

$$\left[\frac{\delta A_F}{A_F}\right]_\mu \doteq \frac{\left[\frac{\delta\mu}{\mu}\right]}{1 - \mu\beta}, \tag{6}$$

$$\left[\frac{\delta A_F}{A_F}\right]_\beta \doteq \frac{\mu\beta}{1 - \mu\beta}\left[\frac{\delta\beta}{\beta}\right]. \tag{7}$$

If $\mu\beta \gg 1$, it is seen that μ or the μ-circuit is stabilized by an amount corresponding to the reduction in amplification and the effect of introducing a gain or loss in the μ-circuit is to produce no material change in the overall amplification of the system; the stability of amplification as affected by β or the β-circuit is neither appreciably improved nor degraded since increasing the loss in the β-circuit raises the gain of the amplifier by an amount almost corresponding to the loss introduced and vice-versa. If μ and β are both varied and the variations sufficiently small, the effect is the same as if each were changed separately and the two results then combined.

In certain practical applications of amplifiers it is the change in gain or ammeter or voltmeter reading at the output that is a measure of the stability rather than the complex ratio previously treated. The conditions surrounding gain stability may be examined by considering the absolute value of A_F. This is shown as follows: Let (db) represent the gain in decibels corresponding to A_F. Then

$$(db) = 20 \log_{10} |A_F|,$$

$$\delta(db) \doteq 8.686\left[\frac{\delta|A_F|}{|A_F|}\right]. \tag{8}$$

To get the absolute value of the amplification: Let

$$\mu\beta = |\mu\beta| \angle \Phi, \tag{9}$$

$$|A_F| = \frac{|\mu|}{\sqrt{1 - 2|\mu\beta| \cos \Phi + |\mu\beta|^2}}. \tag{10}$$

The stability of amplification which is proportional to the gain stability is given by:

$$\left| \frac{\delta |A_F|}{|A_F|} \right|_{|\mu|} \doteq \frac{1 - |\mu\beta| \cos \Phi}{|1 - \mu\beta|^2} \left| \frac{\delta |\mu|}{|\mu|} \right|, \tag{11}$$

$$\left| \frac{\delta |A_F|}{|A_F|} \right|_{|\beta|} \doteq \left| \frac{\mu\beta}{1 - \mu\beta} \right| \left| \frac{\cos \Phi - |\mu\beta|}{|1 - \mu\beta|} \right| \left| \frac{\delta |\beta|}{|\beta|} \right|, \tag{12}$$

$$\left| \frac{\delta |A_F|}{|A_F|} \right|_{\Phi} \doteq - \left| \frac{\mu\beta}{1 - \mu\beta} \right| \left| \frac{\sin \Phi}{|1 - \mu\beta|} \right| [\delta\Phi]. \tag{13}$$

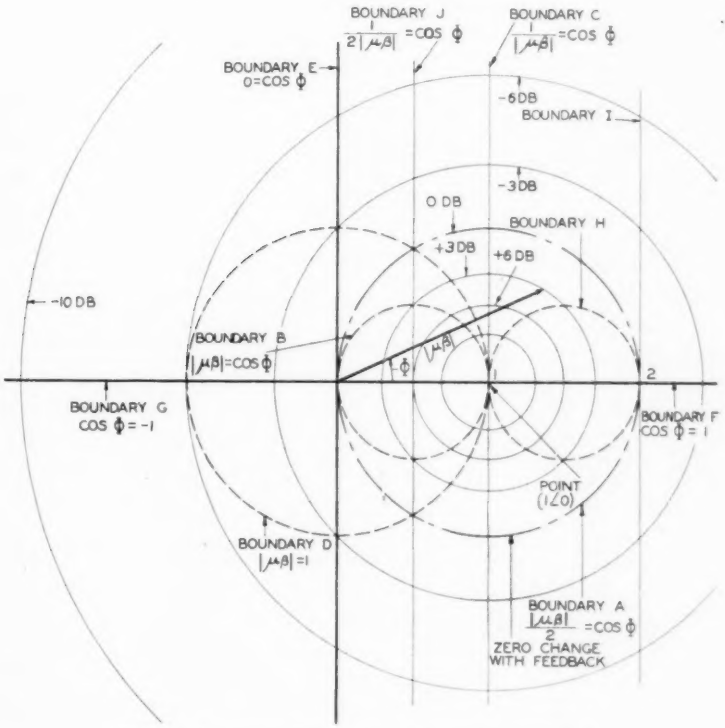


Fig. 3—The vector field of $\mu\beta$. See caption for Fig. 4.

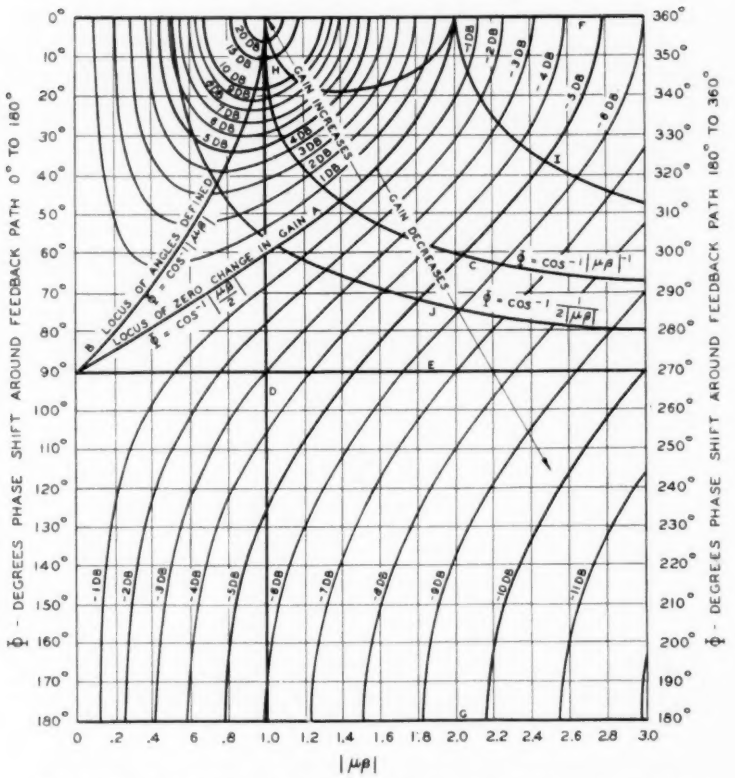


Fig. 4—Phase shift around the feedback path plotted as a function of $|\mu\beta|$, the absolute value of $\mu\beta$.

$\mu\beta$ is a complex quantity which represents the ratio by which the amplifier and feedback (or more generally μ and β) modify a voltage in a single trip around the closed path.

First, there is a set of boundary curves indicated as *A*, *B*, *C*, *D*, *E*, *F*, *G*, *H*, *I*, and *J* which gives either limiting or significent values of $|\mu\beta|$ and Φ.

Secondly, there is a family of curves in which db change in gain due to feedback is the parameter.

Boundaries

A. Conditions in which gain and modulation are unaffected by feedback.

B. Constant amplification ratio against small variations in $|\beta|$.

Constant change in gain, $\frac{1}{|1-\mu\beta|}$, against variations in $|\mu|$ and $|\beta|$.

Stable phase shift through the amplifier against variations in Φ_β.

The boundary on which the stability of amplification is unaffected by feedback.

C. Constant amplification ratio against small variations in $|\mu|$.

Constant phase shift through the amplifier against variations in Φ_μ.

The absolute magnitude of the voltage fed back $\frac{|\mu\beta|}{|1-\mu\beta|}$ is constant against variations in $|\mu|$ and $|\beta|$.

A curious fact to be noted from (11) is that it is possible to choose a value of $\mu\beta$ (namely, $|\mu\beta| = \sec\Phi$) so that the numerator of the right hand side vanishes. This means that the gain stability is perfect, assuming differential variations in $|\mu|$. Referring to Figs. 3 and 4, contour C is the locus of $|\mu\beta| = \sec\Phi$ and it includes all amplifiers whose gain is unaffected by small variations in $|\mu|$. In this way it is even possible to stabilize an amplifier whose feedback is positive, i.e., feedback may be utilized to raise the gain of an amplifier and, at the same time, the gain stability with feedback *need* not be degraded but on the contrary improved. If a similar procedure is followed with an amplifier whose feedback is negative, the gain stability will be theoretically perfect and independent of the reductions in gain due to feedback. Over too wide a frequency band practical difficulties will limit the improvements possible by these methods.

With negative feedback, gain stability is always improved by an amount at least as great as corresponds to the reduction in gain and generally more; with positive feedback, gain stability is never degraded by more than would correspond to the increase in gain and under appropriate conditions, assuming the variations are not too great, is as good as or much better than without feedback. With positive feedback, the variations in μ or β must not be permitted to become sufficiently great to cause the amplifier to sing or give rise to instability as defined in a following section on "Avoiding Singing."

Modulation

To determine the effect of feedback action upon modulation produced in the amplifier circuit, it is convenient to assume that the output of undistorted signal is made the same with and without feedback and that a comparison is then made of the difference in modulation with and without feedback. Therefore, with feedback, the input is changed to $e = e_0(1 - \mu\beta)$ and, referring to equation (2), the output voltage is μe_0, and the generated modulation, $d(E)$, assumes its value without feedback, $d(E_0)$, and $d(E)/(1-\mu\beta)$ becomes $d(E_0)/(1-\mu\beta)$ which is $D_0/(1 - \mu\beta)$. This relationship is approximate because the

---

D. $|\mu\beta| = 1$.

E. $\Phi = 90°$. Improvement in gain stability corresponds to twice db reduction in gain.

F and *G*. Constant amplification ratio against variations in Φ.
Constant phase shift through the amplifier against variations in $|\mu|$ and $|\beta|$.

H. Same properties as *B*.

I. Same properties as *E*.

J. Conditions in which $\dfrac{|\mu|}{|1 - \mu\beta|} = \dfrac{-1}{|\beta|}$ the overall gain is the exact negative inverse of the transmission through the β-circuit.

voltage at the input without feedback is free from distortion and with feedback it is not and, hence, the assumption that the generated modulation is a function *only of the signal output* used in deriving equation (2) is not necessarily justified.

From the relationship $D = D_0/(1 - \mu\beta)$, it is to be concluded that modulation with feedback will be reduced db for db as the effect of feedback action causes an arbitrary db reduction in the gain of the amplifier, i.e., when the feedback is negative. With positive feedback the opposite is true, the modulation being increased by an amount corresponding to the increase in amplification.

If modulation in the β-circuit is a factor, it can be shown that usually in its effect on the output, the modulation level at the output due to non-linearity of the β-circuit is approximately $\mu\beta/(1 - \mu\beta)$ multiplied by the modulation generated in the β-circuit acting alone and without feedback.

Additional Effects

Noise

A criterion of the worth of a reduction in noise is the reduction in signal-to-noise ratio at the output of an amplifier. Assuming that the amount of noise introduced is the same in two systems, for example with and without feedback respectively, and that the signal outputs are the same, a comparison of the signal-to-noise ratios will be affected by the amplification between the place at which the noise enters and the output. Denoting this amplification by a and a_0 respectively, it can be shown that the relation between the two noise ratios is $(a_0/a)(1 - \mu\beta)$. This is called the *noise index*.

If noise is introduced in the power supply circuits of the last tube, $a_0/a = 1$ and the noise index is $(1 - \mu\beta)$. As a result of this relation less expensive power supply filters are possible in the last stage.

Phase Shift, Envelope Delay, Delay Distortion

In the expression $A_F = [\mu/(1 - \mu\beta)]\angle\theta$, θ is the overall phase shift with feedback, and it can be shown that the *phase shift through the amplifier with feedback may be made to approach the phase shift through the β-circuit plus 180 degrees.* The effect of phase shift in the β-circuit is not correspondingly reduced. It will be recalled that in reducing the change in phase shift with frequency, envelope delay, which is the slope of the phase shift with respect to the angular velocity, $\omega = 2\pi f$, also is reduced. The *delay* distortion likewise is reduced because a measure of delay distortion at a particular frequency is the difference between the envelope delay at that frequency and the least envelope delay in the band.

β-Circuit Equalization

Referring to equation (2), the output voltage, E, approaches $-e_0/\beta$ as $1 - \mu\beta \doteq -\mu\beta$ and equals it in absolute value if $\cos\Phi = \dfrac{1}{2|\mu\beta|}$ where $\mu\beta = |\mu\beta| \angle\Phi$. Under these circumstances increasing the loss in the β-circuit one db raises the gain of the amplifier one db and vice-versa, thus giving any gain-frequency characteristic for which a like loss-frequency characteristic can be inserted in the β-circuit. This procedure has been termed β-circuit equalization. It possesses other advantages which cannot be dwelt upon here.

Avoid Singing

Having considered the theory up to this point, experimental evidence was readily acquired to demonstrate that $\mu\beta$ might assume large values,

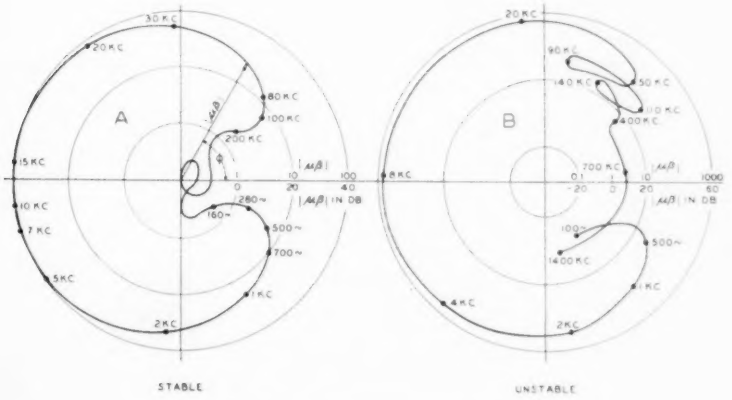


Fig. 5—Measured $\mu\beta$ characteristics of two amplifiers.

for example 10 or 10,000, provided Φ was not at the same time zero. However, one noticeable feature about the field of $\mu\beta$ (Figs. 3 and 4) is that it implies that even though the phase shift is zero and the absolute value of $\mu\beta$ exceeds unity, self-oscillations or singing will not result. This may or may not be true. When the author first thought about this matter he suspected that owing to practical non-linearity, singing would result whenever the gain around the closed loop equalled or exceeded the loss and simultaneously the phase shift was zero, i.e., $\mu\beta = |\mu\beta| + jo \geq 1$. Results of experiments, however, seemed to indicate something more was involved and these matters were described to Mr. H. Nyquist, who developed a more general criterion

for freedom from instability [3] applicable to an amplifier having linear positive constants.

To use this criterion, plot $\mu\beta$ (the modulus and argument vary with frequency) and its complex conjugate in polar coordinates for all values of frequency from 0 to $+\infty$. If the resulting loop or loops do not enclose the point (1, 0) the system will be stable, otherwise not.[3] The envelope of the transient response of a stable amplifier

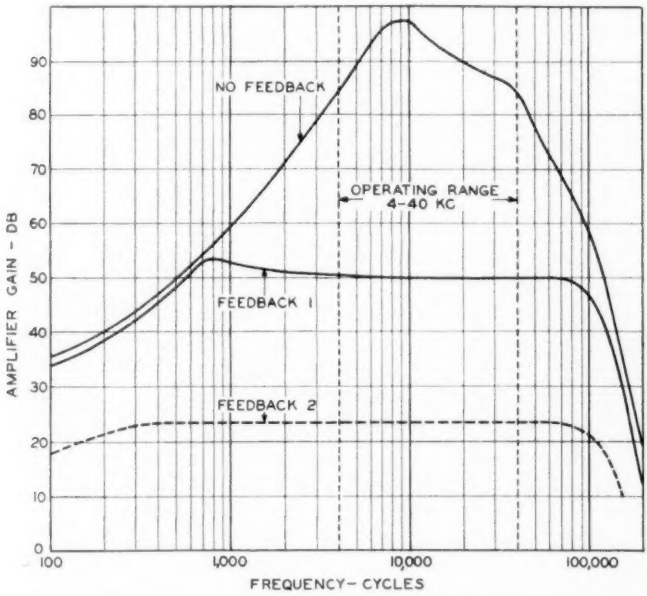


Fig. 6—Gain frequency characteristics with and without feedback of amplifier of Fig. 2.

always dies away exponentially with time; that of an unstable amplifier in all physically realizable cases increases with time. Characteristics *A* and *B* in Fig. 5 are results of measurements on two different amplifiers; the amplifier having $\mu\beta$-characteristic denoted *A* was stable; the other unstable.

The number of stages of amplification that can be used in a single amplifier is not significant except insofar as it affects the question of avoiding singing. Amplifiers with considerable negative feedback

[3] For a complete description of the criterion for stability and instability and exactly what is meant by enclosing the point (1, 0), reference should be made to "Regeneration Theory"—H. Nyquist, *Bell System Technical Journal*, Vol. XI, pp. 126–147, July, 1932.

have been tested where the number of stages ranged from one to five inclusive. In every case the feedback path was from the output of the last tube to the input of the first tube.

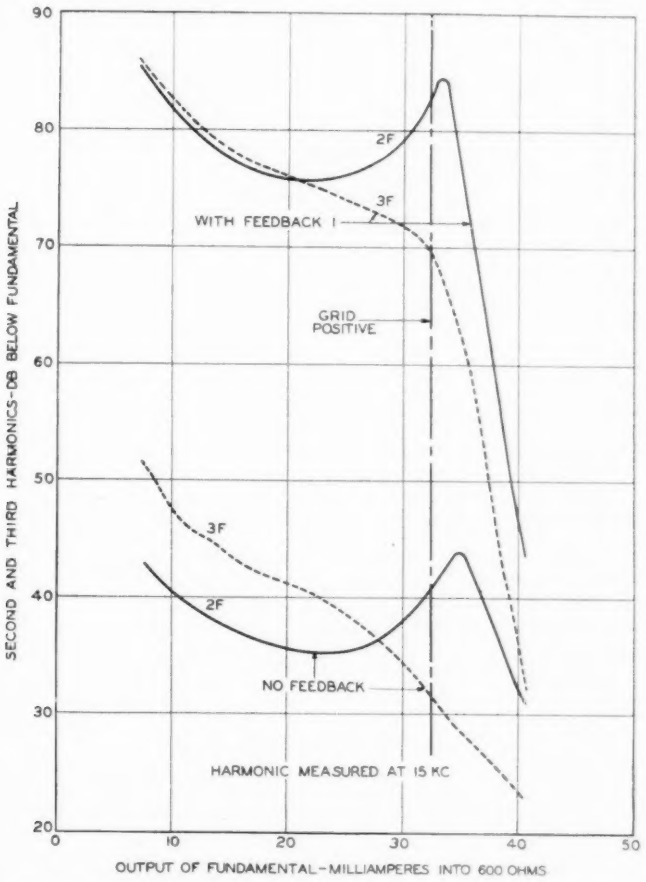


Fig. 7—Modulation characteristics with and without feedback for the amplifier of Fig. 2.

Experimental Results

Figures 6 and 7 show how the gain-frequency and modulation characteristics of the three-stage impedance coupled amplifier of Fig. 2 are improved by negative feedback. In Fig. 7, the improvement in harmonics is not exactly equal to the db reduction in gain. Figure 8

shows measurements on a different amplifier in which harmonics are reduced as negative feedback is increased, db for db over a 65 db range.

That the gain with frequency is practically independent of small variations in $|\mu|$ is shown by Fig. 9. This is a characteristic of the Morristown amplifier described in the paper by Messrs. Clark and Kendall [1] which meets the severe requirements imposed upon a repeater amplifier for use in cable carrier systems. Designed to amplify frequencies from 4 kc

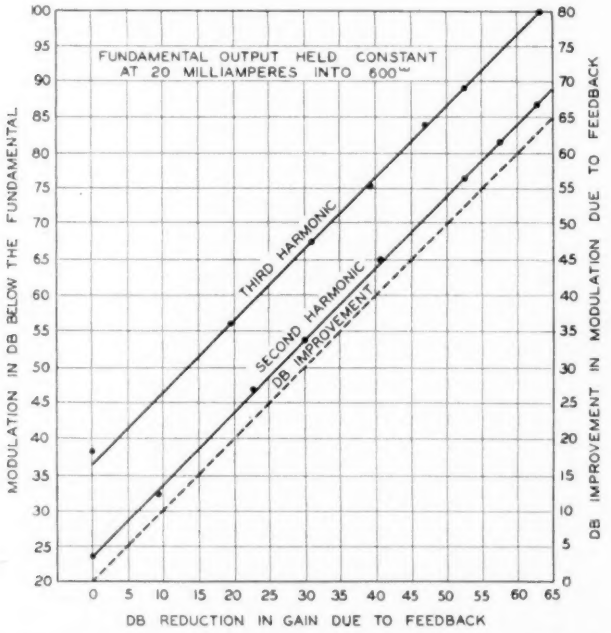


Fig. 8—Improvement of harmonics with feedback. One example of another amplifier in which with 60 db feedback, harmonic currents in the output are only one-thousandth and their energy one-millionth of the values without feedback.

to 40 kc the maximum change in gain due to variations in plate voltage does not exceed 7/10000 db per volt and at 20 kc the change is only 1/20000 db per volt. This illustrates that for small changes in $|\mu|$, the ratio of the stability without feedback to the stability with feedback, called the *stability index*, approaches $|1 - \mu\beta|^2/(1 - |\mu\beta| \cos \Phi)$ and gain stability is improved at least as much as the gain is reduced and usually more and is theoretically perfect if $\cos \Phi = 1/|\mu\beta|$.

[1] Loc. cit.

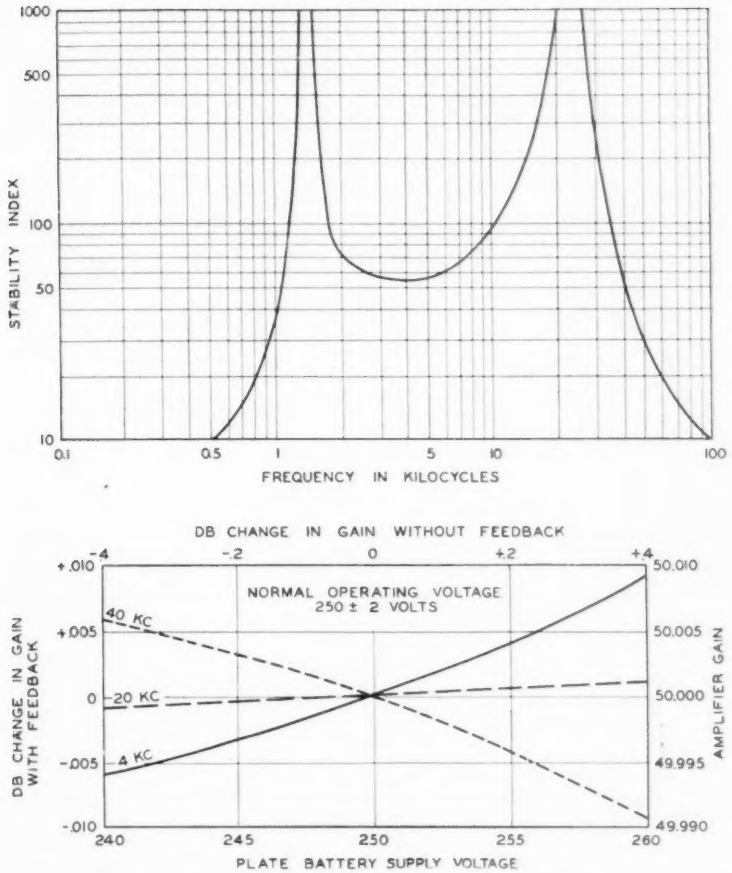


Fig. 9—Representative gain stability of a single amplifier as determined by measuring 69 feedback amplifiers in tandem at Morristown, N. J.

The upper figure shows the absolute value of the stability index. It can be seen that between 20 and 25 kc the improvement in stability is more than 1000 to 1 yet the reduction in gain was less than 35 db.

The lower figure shows change in gain of the feedback amplifier with changes in the plate battery voltage and the corresponding changes in gain without feedback. At some frequencies the change in gain is of the same sign as without feedback and at others it is of opposite sign and it can be seen that near 23 kc the stability must be perfect.

Figure 10 indicates the effectiveness with which the gain of a feedback amplifier can be made independent of variations in input amplitude up to practically the overload point of the amplifier. These measurements were made on a three-stage amplifier designed to work from 3.3 kc to 50 kc.

Figure 11 shows that negative feedback may be used to improve phase shift and reduce delay and delay distortion. These measurements

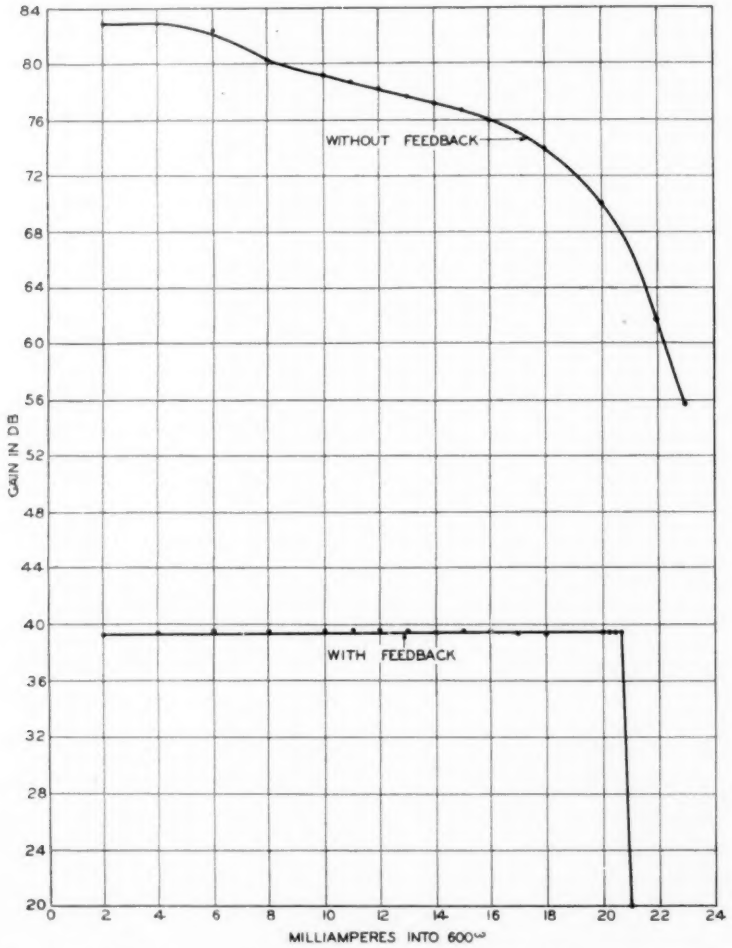


Fig. 10—Gain-load characteristic with and without feedback for a low level amplifier designed to amplify frequencies from 3.5 to 50 kc.

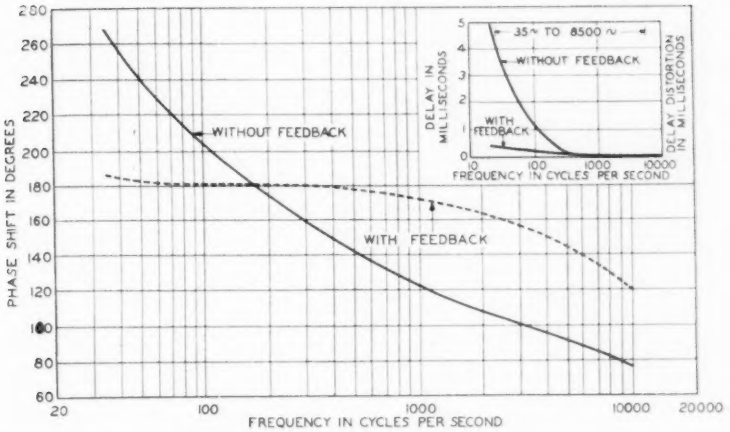


Fig. 11—Phase shift, delay, and delay distortion with and without feedback for a single tube voice frequency amplifier.

were made on an experimental one-tube amplifier, 35–8500 cycles, feeding back around the low side windings of the input and output transformers.

Figure 12 gives the gain-frequency characteristic of an amplifier with and without feedback when in the β-circuit there was an equalizer

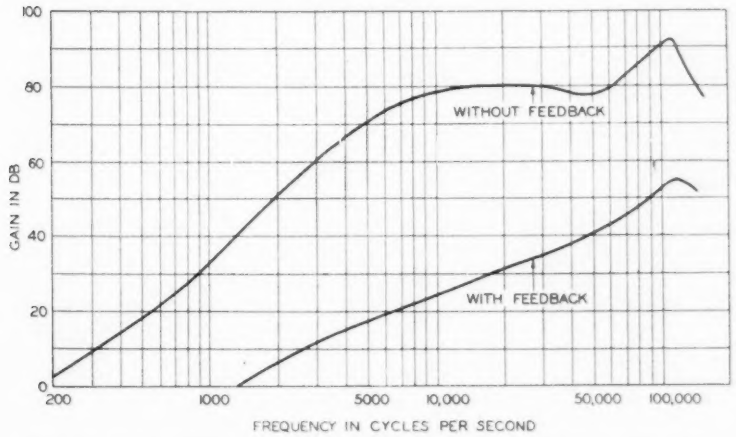


Fig. 12—Gain-frequency characteristic of an amplifier with an equalizer in the β-circuit. This was designed to have a gain frequency characteristic with feedback of the same shape as the loss frequency characteristic of a non-loaded telephone cable.

designed to make the gain-frequency characteristic of the amplifier with feedback of the same shape as the loss-frequency characteristic of a non-loaded telephone cable.

Conclusion

The feedback amplifier dealt with in this paper was developed primarily with requirements in mind for a cable carrier telephone system, involving many amplifiers in tandem with many telephone channels passing through each amplifier. Most of the examples of feedback amplifier performance have naturally been drawn from amplifiers designed for this field of operation. In this field, vacuum tube amplifiers normally possessing good characteristics with respect to stability and freedom from distortion are made to possess superlatively good characteristics by application of the feedback principle.

However, certain types of amplifiers in which economy has been secured by sacrificing performance characteristics, particularly as regards distortion, can be made to possess improved characteristics by the application of feedback. Discussion of these amplifiers is beyond the scope of this paper.

Open-Wire Crosstalk *

By A. G. CHAPMAN

Introduction

THE tendency of communication circuits to crosstalk from one to another was greatly increased by the advent of telephone repeaters and carrier current methods. Telephone repeaters multiplied circuit lengths many times, increased the power applied to the wires, and at the same time made the circuits much more efficient in transmitting crosstalk currents as well as the wanted currents. Carrier current methods added higher ranges of frequency with consequently increased crosstalk coupling. Program transmission service added to the difficulties since circuits for transmitting programs to broadcasting stations must accommodate frequency and volume ranges greater than those required for message telephone circuits.

As these new types of circuits were developed, their application to existing open-wire lines was attended with considerable difficulty from the crosstalk standpoint. Severe restrictions had to be placed on the allocation of pairs of wires for different services in order to keep the crosstalk within tolerable bounds. In many cases the existing lines were retransposed but, nevertheless, there were still important restrictions. While great reduction in crosstalk was obtained by the transposition arrangements the crosstalk reduction was finally limited by unavoidable irregularities in the spacing of the transposition poles and in the spacing of the wires, including differences in wire sag. To further improve matters it was, therefore, necessary to alter the wire configurations so as to reduce the coupling per unit length between the various circuits.

Recently this study of wire configurations has resulted in extensive use of new configurations of open-wire lines in which the two wires of a pair are placed eight inches apart instead of 12 inches, the horizontal separation between wires of different pairs being correspondingly increased. With these eight-inch pairs it has usually been found desirable to discard the time-honored phantoming method of obtaining

* This paper gives a comprehensive discussion of the fundamental principles of crosstalk between open-wire circuits and their application to the transposition design theory and technique which have been developed over a period of years. In this issue of the *Technical Journal* the first half of the paper is published. In the April 1934 issue will be the concluding part, together with an appendix entitled "Calculation of Crosstalk Coefficients."

additional circuits so as to make it possible to obtain a greater number of circuits by more intensive application of carrier current methods.

It is the object of this paper to outline the fundamental principles concerning crosstalk between open-wire circuits and recent developments in transposition design theory and technique which have led to the latest pole line configurations and transposition designs.

To those generally interested in electrical matters it is hoped that this paper will give an insight into the problem of keeping crosstalk in open-wire lines within proper bounds. To those interested in crosstalk it is hoped that the paper will give a useful review of the whole matter and perhaps an insight into the importance of some phenomena which do not seem to be generally appreciated.

The principles set forth in this paper will also be found of considerable interest in connection with problems of control of cable crosstalk, particularly for the high frequencies involved in carrier transmission. It will also be recognized that use is made here of the same general principles as are used in the calculation of effects of impedance irregularities and echoes on repeater operation. These general principles have also been found useful in the development of combinations or arrays of radio antennas of the long horizontal wire type.

The art of crosstalk control in open-wire lines has grown up as a result of the efforts of many workers. The individual contributions are so numerous that it has not been considered practicable in this paper to make individual mention of them except in a few special cases.

General

In the evolution of a satisfactory transposition design technique, complicated electrical actions must be considered and it has been convenient to divide the total crosstalk coupling into various types, all of which may contribute in producing crosstalk between any two circuits in proximity. The first portion of this paper is therefore devoted largely to an examination of the underlying principles and the definition of some of the special terms employed, such as *transverse crosstalk*, *interaction crosstalk*, *reflection crosstalk*, etc. The paper then considers the general effect of transpositions in reducing crosstalk and how this effect depends on the attenuation and phase change accompanying the transmission of communication currents. Consideration is next given to the practical significance of and methods for determining the *crosstalk coefficients* which are used in calculating the crosstalk in a short part of a parallel between two currents. The matter of *type unbalances* inherent in different arrangements of trans-

positions and used in working from short lengths to long lengths is discussed at length. The next section of the paper is devoted to the effect of constructional irregularities caused by pole spacing, wire sag, "drop bracket" transpositions, etc. Various "non-inductive" wire arrangements are considered. The paper closes with a general discussion of practical transposition design methods based on the principles previously disclosed.

Underlying Principles

The discussion under this heading will cover the general causes of crosstalk coupling between open-wire circuits and the general types into which it is convenient to divide the crosstalk effect. The usual measures of crosstalk coupling will also be discussed.

Causes and Types of Crosstalk

The crosstalk coupling between open-wire pairs is due almost entirely to the external electric and magnetic fields of the disturbing circuit. If these fields were in some way annulled there would remain the possibility of resistance coupling between the pairs because of leakage from one circuit to the other by way of the crossarms and insulators, tree branches, etc. This leakage effect is minor in a well-maintained line. It enters as a factor in the design of open-wire transpositions only in so far as the attenuation of the circuits is affected which indirectly affects the crosstalk.

Figure 1 indicates cross-sections of two pairs of wires designated as 1–2 and 3–4. If pair 1–2 existed alone and if the two wires were similar, a voltage impressed at one end of the circuit would result in

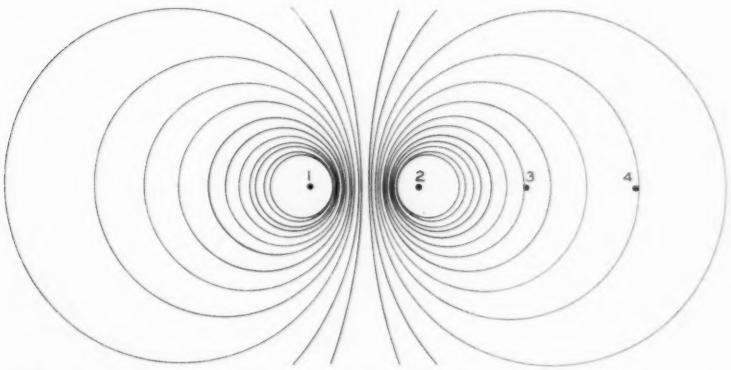


Fig. 1—Magnetic field produced by equal and opposite currents in wires 1 and 2.

equal and opposite currents at any point. These currents would produce a magnetic field as indicated on the figure. If circuit 3–4 parallels 1–2 a certain amount of this magnetic flux would thread between wires 3 and 4 and induce a voltage in circuit 3–4 which would result in a crosstalk current in this circuit. This induced voltage is, of course, due to the difference between the two magnetic fields set up by the opposite directional currents in wires 1 and 2. Since wires 1 and 2 are not very far apart, the resultant field is much weaker than if transmission over wire 1 with ground return were attempted. It is important, therefore, that the wires of a circuit be placed as close together as practicable and that these wires be similar in material and gauge in order to keep the currents practically equal and opposite.

Equal and opposite charges accompany the equal and opposite currents in wires 1 and 2. The equipotential lines of the resultant electric field set up by the two charges are also indicated by Fig. 1. This field will cause different potentials at the surfaces of wires 3 and 4 and this potential difference will cause a crosstalk current in circuit 3–4. As in the case of magnetic induction this current may be minimized by close spacing and electrical similarity between the two wires of a pair.

Calculations of crosstalk coupling must, in general, consider both the electric and magnetic components of the electromagnetic field of the disturbing circuit.

The exact computation of crosstalk coupling between communication circuits is very complex.[1] Approximate computations are sufficient for transposition design. In such computations, it is convenient to divide the total coupling into components of several general types. In calculations of coupling of these types it is assumed that the two wires of a circuit are similar in material and gauge. If there is any slight dissimilarity, such as extra resistance in one wire due to a poor joint, the effect on the crosstalk may be computed separately. The general types of crosstalk coupling are:

1. Transverse crosstalk coupling.
   1*a*. Direct.
   1*b*. Indirect.
2. Interaction crosstalk coupling.

A multi-wire pole line involves many circuits all mutually coupled. In explaining the above terms, it is convenient to start with the simple conception of but two paralleling coupled circuits; Fig. 2A

[1] The general mathematical theory is given in the Carson-Hoyt paper listed under "Bibliography."

indicates such a parallel. In calculating the crosstalk coupling between a terminal of circuit a and a terminal of circuit b, the parallel may be divided into a series of thin transverse slices. One such slice of thickness d is indicated on the figure. The coupling in each slice is calculated and, then, the total coupling between circuit terminals due to all the slices.

In Fig. 2A circuit a is considered to be the disturber and to be energized at the left-hand end. In the single slice indicated, a transmission current will be propagated along circuit a and will cause crosstalk currents in circuit b at both ends of the slice. In this slice, therefore, the left-hand end of circuit a may be considered to be coupled to the two ends of circuit b through the transmission paths n_{ab} and f_{ab}. The path n_{ab} is called the near-end crosstalk coupling and the path f_{ab} is called the far-end crosstalk coupling.

The presence of a tertiary circuit, such as c of Fig. 2B, changes both the near-end and the far-end coupling between a and b in the transverse slice. In addition to the direct couplings n_{ab} and f_{ab} there are indirect couplings n_{acb} and f_{acb} by way of circuit c.

The *transverse crosstalk coupling* between a terminal of a disturbing circuit and a terminal of a disturbed circuit is defined as the coupling between these points due to all the small couplings in all the thin transverse slices including indirect couplings in each slice by way of other circuits. (There are also indirect couplings involving more than one slice and these are not included in the transverse crosstalk coupling.)

In computations of transverse crosstalk coupling it is convenient to distinguish between the direct and indirect components. The direct component considers only the currents and charges in the disturbing circuit while the indirect component takes account of certain charges in tertiary circuits resulting from transmission over the disturbing circuit. The tertiary circuits may be circuits used for transmission purposes or any other circuits which can be made up of combinations of wires on the line or of these wires and ground. If there are only two pairs on the line as in Fig. 2A there are still tertiary circuits, namely, the "phantom" circuit consisting of pair a as one side of the circuit and pair b as the return and the "ghost" circuit consisting of all four wires with ground return. In a multi-wire line many of the tertiary circuits involve the wires of the disturbing circuit. If these tertiary circuits did not exist the currents at any point in the two wires of the disturbing circuit would be equal and opposite. The presence of the tertiary circuits makes these currents unequal and it is convenient to divide the actual currents into two components, i.e.,

equal and opposite or "balanced" currents in the two wires of the disturbing circuit and equal currents in phase in the two wires. The latter may be called "tertiary circuit" currents. The charges on the two wires of the disturbing circuit may be similarly divided into components.

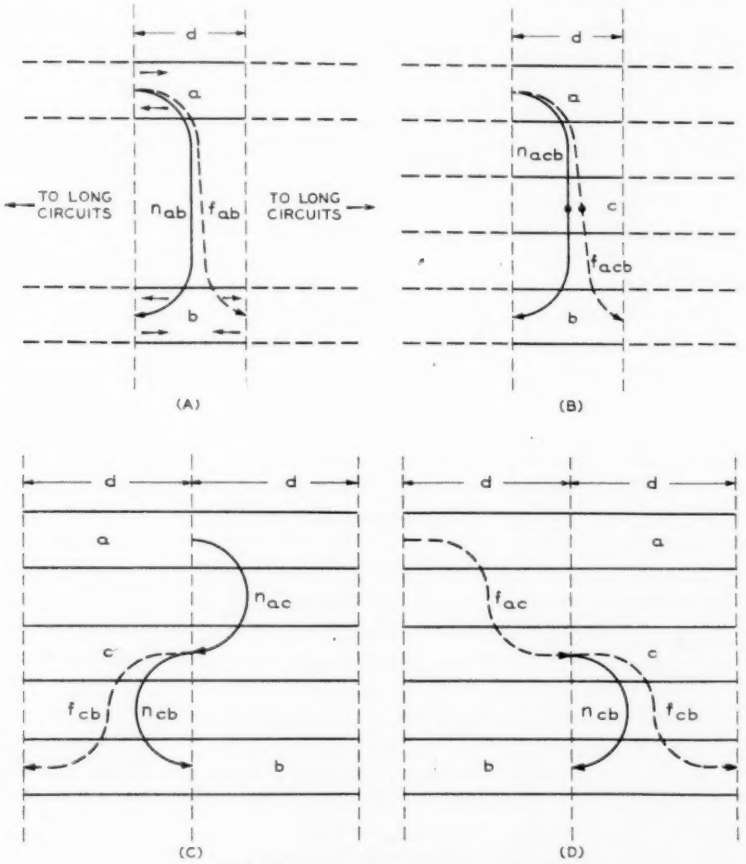


Fig. 2—Transverse and interaction crosstalk.

The *direct* component of the transverse crosstalk coupling is defined as that part which is due to balanced charges and currents in the disturbing circuit.[2] The *indirect* component is defined as that part

[2] "Direct" is here used in a different sense from that used in connection with the term "direct capacity unbalance" which was originated by Dr. G. A. Campbell and has been much used in discussions of cable crosstalk.

which is due to *charges* on tertiary circuits which arise within any thin transverse slice due to coupling with the disturbing circuit in that same slice. This coupling, in any slice, causes currents as well as charges in the tertiary circuit in that slice, but, as discussed in detail in Appendix A, the effect of these currents in producing crosstalk currents in the disturbed circuit is small compared with the effect of the charges. The currents and charges in the tertiary circuits in any thin slice due to the coupling with the disturbing circuit in that same slice may be but a small part of the total currents and charges in the slice. The total values are due to couplings of the tertiary circuits with the disturbing circuit in all the slices. When the total values are considered currents as well as charges in the tertiary circuit may be important in causing crosstalk currents in the disturbed circuit. To consider the total currents and charges in the tertiary circuits it is necessary to take account of both the interaction crosstalk coupling and the transverse crosstalk coupling between disturbing and disturbed circuits.

The nature of interaction crosstalk coupling is indicated by Figs. 2C and 2D which indicate two successive thin transverse slices of width d in a parallel between two circuits a and b and the typical tertiary circuit c. Assuming transmission from left to right on circuit a in Fig. 2C this circuit is coupled with c in the right-hand slice by the near-end crosstalk coupling indicated by n_{ac}. This coupling causes transmission of crosstalk current (and charge) into the left-hand part of circuit c which has both near-end and far-end crosstalk coupling to circuit b. Consideration of these two successive transverse slices, therefore, introduces the two compound couplings $n_{ac}n_{cb}$ and $n_{ac}f_{cb}$. There are two more of these compound couplings as indicated by Fig. 2D. There is a far-end crosstalk coupling between circuits a and c in the left-hand slice which combines with both near-end and far-end couplings in the right-hand slice. The compound types of crosstalk of Fig. 2C and 2D are called interaction crosstalk since the various slices interact on each other in producing indirect couplings. The *interaction crosstalk coupling* between a terminal of a disturbing circuit and a terminal of a disturbed circuit is defined as the coupling between these points due to the indirect couplings involving *all possible combinations* of different thin transverse slices.

The distinction between indirect transverse crosstalk and interaction crosstalk is that the former takes account of the effect of indirect crosstalk from disturbing to tertiary to disturbed circuit in a single thin transverse slice while the latter involves indirect crosstalk from primary circuit to tertiary circuit in one slice, transmission along the

tertiary circuit into another slice and then crosstalk from tertiary circuit to disturbed circuit.

The notion that there is only transverse crosstalk within any one "thin slice" implies that the slice thickness corresponds to a distance along the line of only infinitesimal length. If this distance were finite it would correspond to a series of "thin slices" having interaction crosstalk between them. Practically, however, if the distance along the line corresponds to a line angle of five degrees or less, the interaction crosstalk in this length is small compared with the transverse crosstalk. A five degree line angle corresponds to a length of about .1 mile at 25 kilocycles, .05 mile at 50 kilocycles, etc. A transposed line is divided into short lengths or *segments* by the transposition poles and the line angle of these segments is ordinarily less than five degrees at the highest frequency for which the transposition system is suitable. Therefore, the crosstalk coupling between such transposed circuits may be computed on the basis of transverse crosstalk within any segment and interaction crosstalk between any two segments.

As shown by Fig. 2 the interaction effect involves the four compound couplings:

$$n_{ac}n_{cb}, \qquad n_{ac}f_{cb}, \qquad f_{ac}n_{cb}, \qquad f_{ac}f_{cb}.$$

The near-end crosstalk couplings n_{ac} and n_{cb} of Fig. 2 are usually much larger than the far-end couplings f_{ac} and f_{cb}. The reason for this, as discussed in Appendix A, is that the electric and magnetic fields of the disturbing circuit tend to aid each other in producing near-end crosstalk coupling such as n_{ac}, and to oppose each other in the case of far-end coupling such as f_{ac}. For this reason the compound coupling $n_{ac}n_{cb}$ is the most important and is usually the only compound coupling which requires consideration in transposition design. Since the path $n_{ac}n_{cb}$ results in a crosstalk current at the far end of the disturbed circuit, it is in connection with far-end crosstalk between long circuits that this matter of interaction crosstalk is important. Far-end rather than near-end crosstalk coupling is controlling in connection with open-wire carrier frequency systems for the reasons explained below.

Figure 3A indicates very schematically two one-way carrier frequency channels routed over two long paralleling open-wire pairs. The boxes at the end indicate the repeaters or terminal apparatus and the arrows on these boxes the direction of transmission of this apparatus. Transmission from the left on pair a results in near-end and far-end crosstalk into pair b, as indicated by the couplings n_{ab} and f_{ab}. The near-end crosstalk current cannot pass to the input of the terminal

apparatus since the latter is a one-way device. In practice, to obtain two-way circuits each of these one-way channels is associated with another one-way channel transmitting in the opposite direction over the same pair of wires. These return channels utilize a different band of carrier frequencies and the near-end crosstalk current is largely excluded from this frequency band by selective filters. The far-end crosstalk is, therefore, the sole consideration with such a carrier system. Use is not made of the same carrier frequencies in both directions on a toll line largely because of difficulties in controlling the near-end crosstalk.

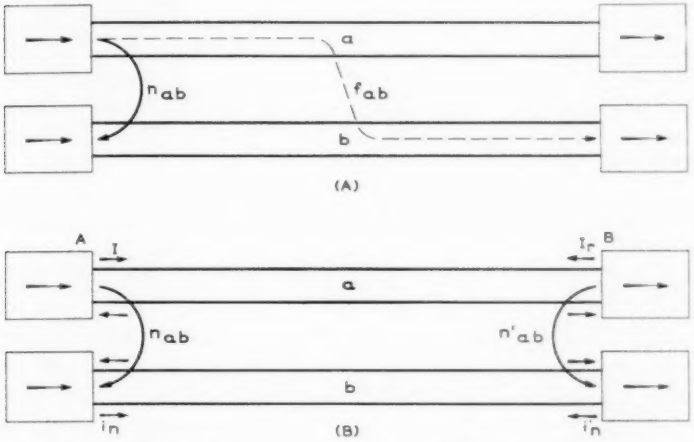


Fig. 3—Crosstalk between two one-way carrier frequency channels.

In connection with the arrangement of Fig. 3A, there is a type of crosstalk of considerable practical importance known as "*reflection crosstalk.*" The theory of this is indicated by Fig. 3B which shows the same two one-way carrier channels. Transmission from left to right on circuit *a* is assumed. When the transmission current I arrives at point B, a certain portion of it will be reflected if there is any deviation of the input impedance of the terminal apparatus from the characteristic impedance of circuit *a*. This reflected current I_r causes a near-end crosstalk current i_n' at point B in the disturbed circuit. Similarly, a part of the near-end crosstalk current i_n at point A in the disturbed circuit may be reflected and transmitted to point B. Therefore, two additional crosstalk currents may result from these two reflections and such currents can enter the terminal apparatus at B and pass through to the output of this apparatus.

For like circuits, like impedance mismatches and like near-end crosstalk couplings at the two ends of the line, these two additional far-end crosstalk currents are of equal importance. Similar reflection effects will occur at any intermediate points in the lines having impedance irregularities. Since the far-end crosstalk coupling can be much more readily reduced by transpositions than the near-end crosstalk coupling this reflection crosstalk effect is important in practice. It is, therefore, necessary to carefully design the terminal and intermediate apparatus and cables to minimize impedance mismatches as far as practicable.

In calculation of crosstalk coupling it is ordinarily assumed that the two wires of a circuit are electrically similar or "balanced" (except as regards crosstalk from other wires). This is substantially true in practice except for accidental deviations, such as resistance differences due to poor joints and leakage differences due to cracked insulators, foliage, etc. Resistance differences may be of considerable practical importance and are said to cause *resistance unbalance crosstalk*. The following discussion indicates the general nature of this effect.

As discussed in connection with Fig. 1, the external field of the disturbing circuit is minimized by the opposing effects of substantially equal and opposite currents or charges in the two wires of the circuit. The two wires may be considered as two separate circuits, each having its return in the ground. At any point in the line these two wires would normally have practically equal and opposite voltages with respect to ground. These voltages would normally cause almost equal and opposite currents in the two wires. If the resistance of one wire is increased due to a bad joint, the current in that wire is reduced and the currents in the two wires are no longer equal and opposite. The external field of the two wires and the resulting voltage induced in the disturbed circuit are, therefore, altered. If this voltage had previously been practically cancelled out by means of transpositions, the alteration in the field would increase the crosstalk current at the terminal of the disturbed circuit.

A resistance unbalance in the disturbed circuit will have a similar effect as indicated by Fig. 4A. This figure shows a short length d of two long paralleling circuits. Equal and opposite transmission currents in the disturbing circuit 1–2 are indicated by I. Equal crosstalk currents in the two wires of the disturbed circuit 3–4 at one end of the short length are indicated by i. It is assumed that these crosstalk currents have been made substantially equal by transpositions in other parts of the line. Since the currents in wires 3 and 4 are equal and in the same direction, there will be no current in a receiver connected

at the terminal of the line between these wires. If, however, one wire has a bad joint, the two crosstalk currents become unequal and there will be a current in such a receiver.

Resistance unbalance crosstalk is of particular importance if two pairs are used to create a phantom circuit in order to obtain three transmission circuits from the four wires. The distribution of the phantom transmission current I_p in a short length of the two pairs is indicated by Fig. 4B. Ideally, half the phantom current flows in each of the four wires. The two currents in wires 1 and 2 are then equal and in the same direction and there will be no current in terminal apparatus connected between wires 1 and 2. In other words, transmission over the phantom circuit results in no crosstalk in the side circuit 1–2. The same may be said of side circuit 3–4. A bad joint in any wire, such as 3, makes the two currents in wires 3 and 4 unequal and results in a current in the side circuit 3–4.

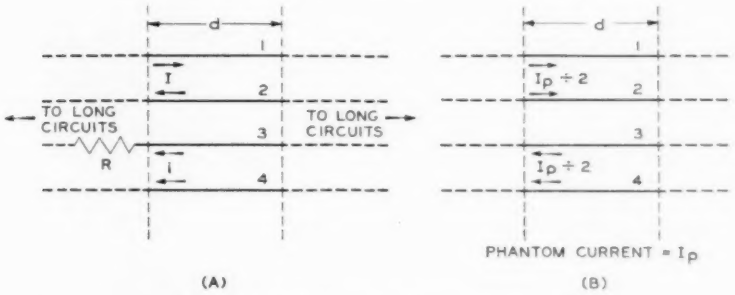


Fig. 4—Effect of resistance unbalance on crosstalk.

The phantom-to-side crosstalk effect of resistance unbalance is much more severe than the effect on crosstalk between two side circuits or two non-phantomed circuits. The reason for this is evident from Figs. 4B and 4A. In Fig. 4B the entire transmission current of the disturbing phantom circuit normally flows in the two wires of the disturbed side circuit and if a resistance unbalance causes a small percentage difference in the currents in these two wires objectionable crosstalk results. In Fig. 4A only crosstalk currents flow in wires 3 and 4 and a much larger percentage difference between these small currents can be tolerated.

In designing and operating phantom circuits, it is necessary to exercise great care to minimize any dissimilarity between the two wires of a side circuit, in order to avoid crosstalk from a phantom to its side circuit or vice versa. Otherwise, the problem of crosstalk between

a phantom circuit and some other circuit is generally similar to the problem of crosstalk between two pairs. In other words, the discussion of transverse, interaction and reflection crosstalk is applicable.

Measures of Crosstalk Coupling

In designing transposition systems, the usual measure of the coupling effect between two open-wire circuits is the ratio of current at the output terminal of the disturbed circuit to current at the input terminal of the disturbing circuit. For circuits of different characteristic impedances this current ratio must be corrected for the difference in impedance. The corrected current ratio is the square root of the corresponding power ratio.

The current ratio is ordinarily very small and for convenience is multiplied by 1,000,000 and called the crosstalk coupling or, in brief, the crosstalk. This usage will be followed from this point in this paper. For example, crosstalk of 1000 units means a current ratio of .001. Crosstalk may also be expressed as the transmission loss in db corresponding to the current ratio. A ratio of .001 means a transmission loss of 60 db corresponding to 1000 crosstalk units.

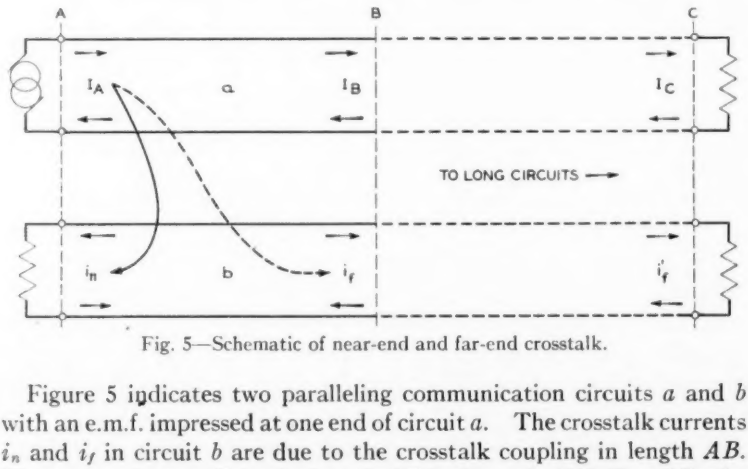


Fig. 5—Schematic of near-end and far-end crosstalk.

Figure 5 indicates two paralleling communication circuits a and b with an e.m.f. impressed at one end of circuit a. The crosstalk currents i_n and i_f in circuit b are due to the crosstalk coupling in length AB. The *near-end crosstalk* in the length AB is the ratio $10^6 i_n/I_A$, while the *far-end crosstalk* is $10^6 i_f/I_A$. The ratio $10^6 i_f/I_B$ has been called the "output-to-output" or "measured" crosstalk. This ratio is a convenient measure of far-end crosstalk between parts of similar circuits because it is related in a simple way to the far-end crosstalk between the terminals of the complete circuits. The following discussion explains this relation.

Both of the currents I_B and i_f will be propagated to point C. They will be attenuated or amplified alike if the circuits are similar and their ratio will be unchanged. The output-to-output crosstalk at C due to the length AB will, therefore, be the same as that determined for point B. In other words $10^6 i_f'/I_C$ will equal $10^6 i_f'/I_B$. The far-end crosstalk between the terminals A and C, due to length AB, will be $10^6 i_f'/I_A$. This differs from the output-to-output crosstalk at C in that the reference current is I_A instead of I_C. The part of the far-end crosstalk between A and C due to AB is, therefore, obtained from the output-to-output crosstalk at B by simply multiplying by the attenuation ratio I_C/I_A. If the output-to-output crosstalk is expressed as a loss in decibels, the far-end crosstalk is obtained by adding the net loss of the complete circuit between A and C.

Effects of Transpositions

The effects of transpositions on both the transmission currents and the crosstalk currents will now be discussed in a general way. The general method of computing the crosstalk between circuits without constructional irregularities and transposed in any manner will also be outlined.

General Principles

If there is only one circuit on a pole line, and this is balanced and free from irregularities, the communication currents will be propagated along this circuit according to the simple exponential law. If a current is propagated from the start of the circuit to some other point at a distance L, the magnitude of the current will be reduced by the attenuation factor $\epsilon^{-\alpha L}$ and the phase of the current will be retarded by the angle βL where α is the attenuation constant and β is the phase change constant.

If there are a number of circuits on a pole line this simple law of propagation may be altered due to crosstalk into surrounding circuits. This is illustrated by the curves, Fig. 6, which indicate the relation between observed output-to-input current ratio and frequency for two different circuits, each about 300 miles long and having 165-mil copper wires. The number of decibels corresponding to the current ratio is plotted rather than the ratio itself. For the simple law of propagation such curves would show the number of decibels increasing smoothly with frequency due to increasing losses in the line wires and insulators. The upper curve is for a circuit too infrequently transposed for the frequency range covered and the current ratio is abnormally small at particular frequencies. The corresponding number of decibels is abnormally large. The lower curve is for a circuit much more fre-

quently transposed and its current ratios practically follow the simple propagation law mentioned above over the frequency range shown.

Even though a circuit is very frequently transposed, its propagation constant is slightly affected by the presence of other circuits on the line. This may be explained by consideration of Figs. 2B and 7. As previously explained, Fig. 2B indicates the indirect transverse crosstalk by way of a tertiary circuit in one thin transverse slice of a parallel

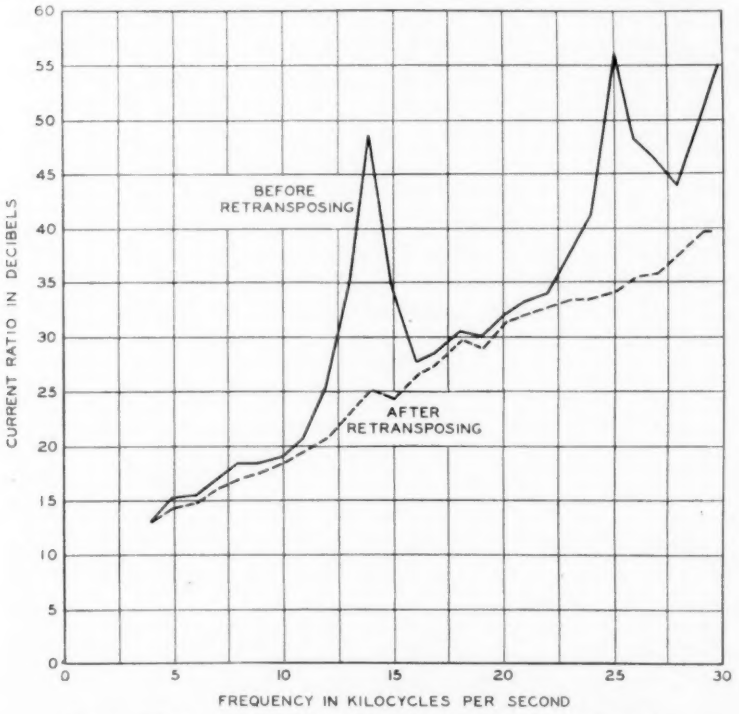


Fig. 6—Effect of transpositions on attenuation of an open-wire pair.

between two long circuits *a* and *b*. The circuit *c* has currents and charges due to crosstalk from the disturbing circuit *a*. These currents and charges not only alter the crosstalk currents in circuit *b* but also react to change the transmission current in circuit *a*. Since circuits *a* and *c* are loosely coupled, this reaction effect could usually be estimated with sufficient accuracy by calculating the crosstalk from *a* to *c* and back again and neglecting the further reactions of the change in the current in *a* on the current in *c*, etc.

Figure 7 shows the crosstalk paths from *a* to *c* and back again. In this figure, circuit *a* is indicated as two separate circuits for comparison with Fig. 2B. It is assumed that circuit *a* in Fig. 7 is energized at point *A*, the currents I_A and I_B being the currents which would exist at the input and output of the short length *d* if there were no tertiary circuits. The near-end crosstalk path indicated by *n* will cause a small crosstalk current i_n at point *A* in circuit *a*. There will be a crosstalk path similar to *n* in each thin slice of the parallel between *a* and *c*. Each of these paths will transmit a small crosstalk current to point *A* in circuit *a*. The sum of all these crosstalk currents will *increase* the input current I_A and, therefore, the impedance of circuit

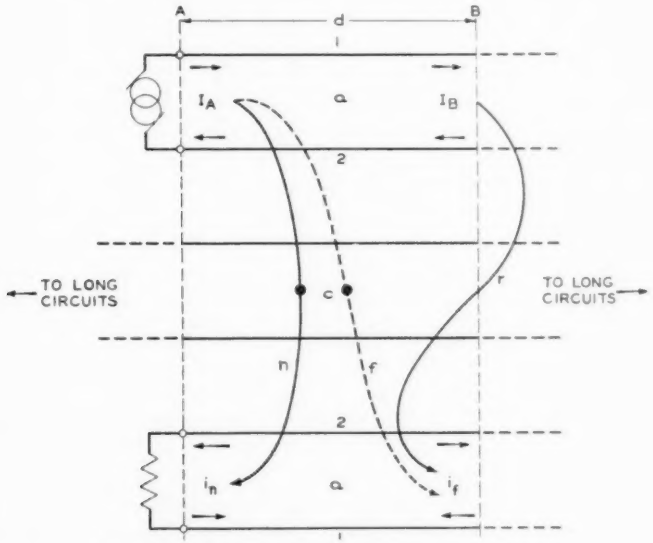


Fig. 7—Effect of circuit c on propagation in circuit a.

a is lowered. Thin slices remote from the sending end will contribute little to this effect, since the crosstalk currents from such slices will be attenuated to negligible proportions. A long circuit on a multi-wire line will, therefore, have a definite sending-end impedance slightly lower than that for one circuit alone on the line.

Figure 7 also indicates a far-end crosstalk path *f* which produces a crosstalk current i_f at point *B* in circuit *a*. This *reduces* the transmission current I_B at this point and, therefore, increases the attenuation constant of the circuit. For calculations of both the circuit

impedance and attenuation, the effect of surrounding circuits is taken care of in practice by using a capacity per unit length slightly higher than the value which would exist with only one circuit on the line. The proper capacity to use is determined in practice by measurements on a short length of a multi-wire line.

The effect on the propagation constant of the transverse crosstalk paths indicated by *n* and *f* of Fig. 7 cannot be suppressed by transpositions. As explained later, if the two circuits marked *a* were actually different circuits, the effect could be largely suppressed by transposing one circuit at certain points and leaving the other circuit untransposed at these points. Since the disturbing and disturbed circuits indicated by Fig. 7 are actually the same circuit, they must be transposed at the same points and, therefore, the transverse crosstalk effect cannot be suppressed by frequent transpositions.

Figure 7 also shows a crosstalk path marked *r*. This is one of the possible interaction crosstalk paths. The effect of such paths on the impedance and attenuation of the circuit may be largely suppressed by suitable transpositions. The difference between the two curves of Fig. 6 is due to lack of this suppression in the case of the upper curve.

Such an extreme effect of crosstalk reacting back into the primary or initiating transmission circuit and thus affecting direct transmission is seldom important in practical transposition design. A marked reaction on the primary circuit would necessitate such large crosstalk currents in neighboring communication circuits as to make them unfit for communication service at the frequency transmitted over the primary circuit. Therefore, it is only when the neighboring circuits are not to be used at this frequency that transposition design to control simply the direct transmission becomes of practical importance. When many circuits on a line are used for carrier operation, the crosstalk currents must be made so weak (by transpositions, physical separation of circuits, etc.) that their reactions back into the primary circuits are very small.

The effect of transpositions on crosstalk from one circuit into another *different* circuit will now be considered. The discussion of the control of this effect is the main object of this paper.

Figure 8A shows a short segment of a parallel between two long circuits and a near-end crosstalk coupling marked *n*. The segment could be divided into a series of thin slices and theoretically there would be interaction crosstalk between different slices. The segment length is, however, assumed to be short enough to neglect interaction crosstalk. The coupling *n* is, therefore, due either to direct or indirect transverse crosstalk in the short segment or to both of these types of

crosstalk. If circuit *a* is energized from the left, a near-end crosstalk current i_n results at point *A* in circuit *b*.

If two successive short segments are considered, as indicated by Fig. 8B, there will be a near-end crosstalk coupling *n* in each segment and each of these couplings will result in a crosstalk current at point *A*

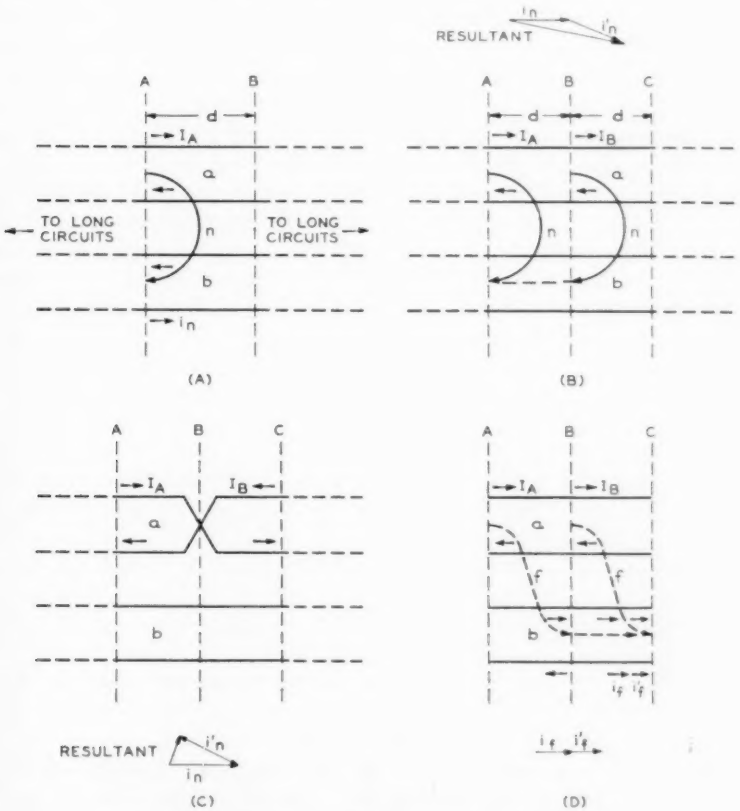


Fig. 8—Effect of transpositions on transverse crosstalk.

of circuit *b*. This is indicated by the vector diagram over the figure, where i_n indicates the crosstalk current due to the segment *AB* and i_n' indicates the crosstalk current at *A* due to *BC*. The latter current is slightly smaller and slightly retarded in phase with respect to i_n because in order for i_n' to appear at point *A*, the transmission current I_A must be propagated a distance *d* and the resulting crosstalk current

at B must also be propagated a distance d in order to reach A. As indicated by this vector diagram the total crosstalk current due to the two short segments is a little less than the arithmetic sum of the individual crosstalk currents.

Figure 8C is like Fig. 8B except that a transposition is inserted in the middle of circuit a at point B. This reverses the phase of the transmission current at the right of B and also reverses any crosstalk current due to current in circuit a between B and C. As a result the crosstalk current i_n' of Fig. 8B is reversed and the resultant of the two crosstalk currents is very much reduced as indicated by the vector diagram of Fig. 8C. The angle between i_n and i_n' is proportional to the length $2d$ which equals AC. The tendency for the two currents to cancel may, therefore, be increased by reducing the length AC which, in a long line, would mean increasing the number of transpositions.

Figure 8D is like Fig. 8B except that the far-end transverse crosstalk coupling f in each of the two short segments is considered. The coupling in the left-hand segment results in a crosstalk current at point B of circuit b, which is propagated to point C as indicated by i_f. The far-end crosstalk coupling in the right-hand segment produces a crosstalk current i_f' at point C. Since the total propagation distance is from A to C for both of these crosstalk currents, they must be equal in magnitude and in phase if circuits a and b are similar. This is indicated by the vector diagram of Fig. 8D. A transposition at point B in either circuit would reverse one of these crosstalk currents and, therefore, the resultant crosstalk current would be nil.

From consideration of Figs. 8C and 8D, it may be seen that if both circuits were transposed at point B, the sum of the crosstalk currents for the two segments would be the same as if neither circuit were transposed. Transposing one circuit reverses the phase of one of the component crosstalk currents, but if the second circuit is also transposed the original phase relations between the two currents are restored.

The foregoing discussion applies only to transverse crosstalk as discussed in connection with Fig. 2. When interaction crosstalk must be considered, a different principle is involved.

In connection with Fig. 8D, it was shown that the transverse far-end crosstalk between similar circuits could be readily annulled by transposing one of the circuits at the center of their paralleling length. Far-end crosstalk of the interaction type is not so readily annulled. The effect of transpositions on this type of crosstalk is indicated by Fig. 9.

This figure shows four short segments in a parallel between two

circuits *a* and *b*, there being an interposed tertiary circuit *c*. Interaction crosstalk involving two near-end crosstalk couplings is considered since this is usually the controlling type. There is an interaction crosstalk path designated *r* between the first two segments as indicated by Fig. 9A. There is a similar path between the third and fourth segments. Each of these paths would produce a far-end crosstalk current in circuit *b* at point *E*. For similar circuits these currents would be equal in magnitude and would add directly. The two currents can be made to cancel by transposing one of the circuits at *C*, the midpoint of the parallel. Such a transposition also cancels the transverse far-end crosstalk in length *AC* against that in length *CE*. There remains, however, the interaction crosstalk between length *CE* and length *AC*.

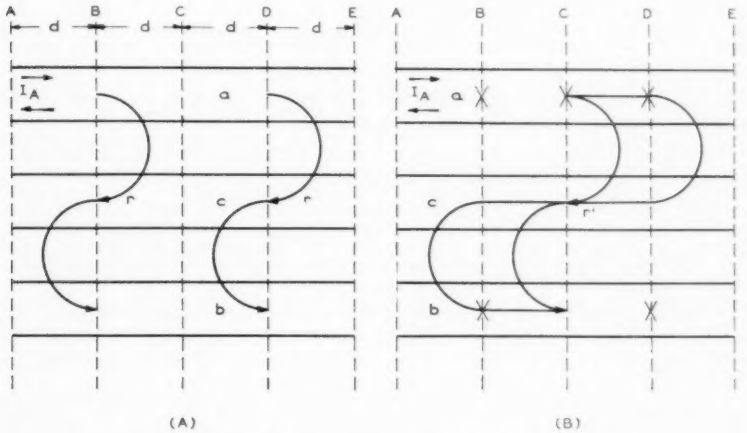


Fig. 9—Effect of transpositions on interaction crosstalk.

Figure 9B shows a transposition at *C* in circuit *a* and also other transpositions whose purpose is to minimize the interaction crosstalk between length *CE* and length *AC*. This crosstalk coupling, designated by *r'*, is a compound effect, depending on the near-end crosstalk between circuit *a* and circuit *c* in length *CE* and the near-end crosstalk between *c* and *b* in length *AC*. The near-end crosstalk coupling between *a* and *c* in length *CE* can be greatly reduced by a transposition in circuit *a* at point *D*, while the crosstalk coupling between *c* and *b* in length *AC* can likewise be reduced by a transposition at point *B* in circuit *b*. The latter two transpositions would not, however, minimize the interaction crosstalk between *CE* and *AC* with circuit *b* as the

disturbing circuit and it is necessary, therefore, to transpose both circuits at points B and D. The addition of these four transpositions does not affect the cancellation of far-end crosstalk in length AC against that in length CE by means of the transposition at C. After the four transpositions are added, length AC is still similar to length CE and the far-end crosstalk currents at E, due to these two lengths, are equal. Therefore, they will cancel when one of them is reversed in phase by the transposition at C.

It may be concluded that, while transposing both circuits at the same points has no effect on transverse crosstalk, it has a large effect on the interaction crosstalk. An experimental illustration is given in Fig. 10. This figure shows frequency plotted against output-to-output

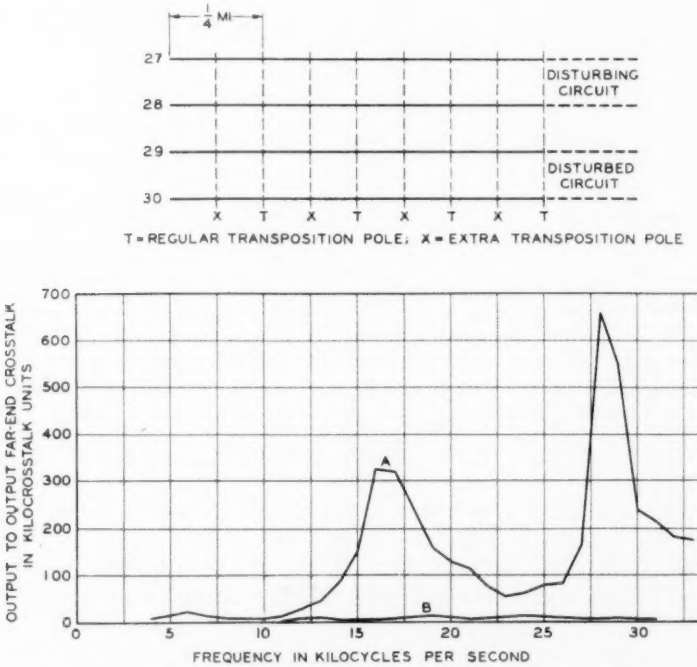


Fig. 10—Effect on far-end crosstalk of extra transpositions in both circuits.

far-end crosstalk between the two side circuits of a phantom group on a 140-mile length of line. The curve marked A is for the two circuits transposed for voice-frequency operation. Curve B is for the two circuits transposed in the same manner except that four transpositions

per mile were added to both circuits at the same points which are indicated by x on Fig. 10. The large effect of these transpositions shows the practical importance of the interaction type of far-end crosstalk.

In connection with Fig. 9B, there arises the question of how far apart the transpositions can be placed without serious crosstalk, in other words, how long is it permissible to make the segment d. If this length is increased the transpositions at B and D become less effective in suppressing the near-end crosstalk between a and c in length CE and between c and b in length AC. The degree to which the interaction crosstalk path r' must be suppressed is, therefore, important in determining the maximum permissible length of d. If d is increased the transposition at C becomes less effective in controlling the near-end crosstalk between a and b and, therefore, the length d also depends on the permissible near-end crosstalk.

It may be noted that transpositions at B and D in but one of the circuits a or b will help to suppress r', but the suppression is less effective than if both circuits are transposed at these points. If a is transposed at B and D the near-end crosstalk between a and c in length CE is reduced but the near-end crosstalk between c and b in length AC is not reduced. The product of these two near-end crosstalk values is greater, therefore, than if they had both been reduced by transposing both circuits at B and D.

Crosstalk Coefficients

The crosstalk between any two long open-wire circuits may be calculated by dividing the parallel into a succession of thin transverse slices and summing up the crosstalk for all these slices. To calculate the crosstalk in any slice it is necessary to know certain "crosstalk coefficients." The discussion below defines these coefficients and describes briefly how they are measured or computed.

Figures 2A and 2B indicate both near-end and far-end crosstalk coupling of both the direct and indirect transverse types in a thin transverse slice. Any of these couplings may be expressed in crosstalk units and the value of the coupling in a short length divided by the length in miles is called the crosstalk per mile. Since, as shown in the previous section, the crosstalk may not increase directly as length, strictly speaking, the crosstalk per mile is the limit of the ratio of coupling to length as the length approaches zero. The crosstalk per mile includes both the direct and indirect types of transverse crosstalk coupling. In the frequency range of interest (i.e., above a few hundred cycles for near-end crosstalk and above a few thousand cycles for

far-end crosstalk) this total transverse coupling varies about directly with the frequency and the crosstalk coefficient commonly used is the *crosstalk per mile per kilocycle.*

If many wires are involved, it is impracticable to determine these coefficients with good accuracy by computation and they are, therefore, derived from measurements. Examples of near-end and far-end coefficients, plotted against frequency, are shown in Fig. 11. The coefficients are for pairs designated 1–2 and 3–4 on the pole head diagram shown on the figure. These coefficients were derived from measurements of the near-end and far-end crosstalk over a range of frequencies. The length of line was about .2 mile and, for the range of frequencies covered, this length is sufficiently short so that interaction crosstalk is negligible and the transverse crosstalk is directly proportional to the length. The coefficients plotted are, therefore, nearly equal to the measured values of crosstalk divided by the length and by the frequency. (A small correction was made at the higher frequencies to allow for deviation of near-end crosstalk from simple proportionality to length and the curves were "smoothed" through the actual points calculated from the measurements.)

In order to obtain the crosstalk coefficients applicable to a short part of a long line, all the wires on the line were terminated in such a manner as to roughly simulate their extension for long distances in both directions, but without crosstalk coupling between the test pairs in such extensions. This is done by terminating each pair at each end with a resistance approximating its characteristic impedance and connecting the midpoint of each resistance to ground through a second resistance. These latter resistances terminate any phantom of two pairs as indicated on Fig. 11 for pairs 1–2 and 3–4. Any circuit with ground return is also terminated by these resistances.

Both of the test pairs are transposed at the midpoint of the line during the measurement. This minimizes the currents reaching the ends of the tertiary circuits and makes even the above approximate termination of the tertiary circuits of little importance.

Figure 11 shows near-end and far-end crosstalk coefficients for three conditions, *A*, *B*, and *C*. The two curves marked *A* show the measured values with all wires terminated and the test pairs transposed as described above.

For curves *B*, only the transposed test pairs were terminated as described above and the other wires were opened at the middle, at the quarter points and at both ends. Since no section of any of these wires connected points of substantially different potential in the field of the disturbing circuit there were practically no currents or charges

in these wires and the crosstalk coefficients for the two test pairs were practically the same as if the other wires had been removed from the line. It will be seen that the crosstalk coefficients for curves *B* are

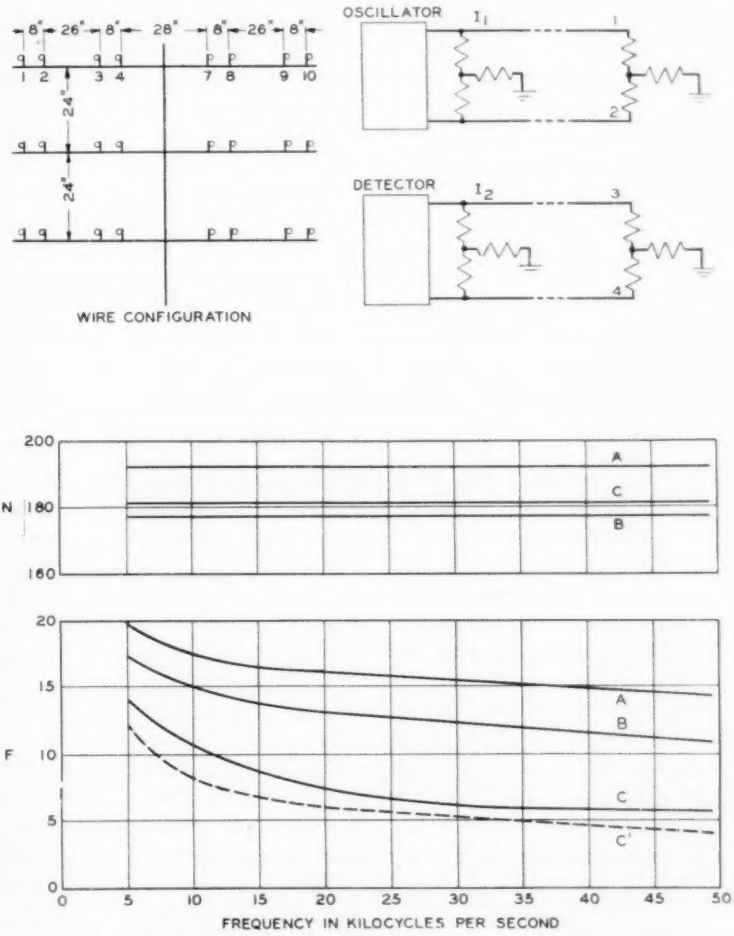


Fig. 11—Near-end and far-end crosstalk coefficients between pairs 1–2 and 3–4.

less than those for curves *A*. The coefficients of curves *B* involve tertiary circuits, however, since there could be crosstalk currents in the phantom of the two test pairs and also in the ghost circuit involving wires 1 to 4 with ground return.

Curves *C* show the coefficients with the test pairs *without* transpositions and terminated at both ends as accurately as practicable, but without the midpoints of these terminations connected to ground to terminate the phantom and ghost circuits. These tertiary circuits were, with this arrangement, prevented from connecting points of substantially different potential and the coefficients of curves *C*, therefore, approach the direct crosstalk coefficients. It is extremely difficult to experimentally determine the direct far-end coefficient. It may be computed, however, and the computed value which assumes perfect terminations and the effect of the phantom completely removed is shown by curve *C'*.

It may be noted that the near-end crosstalk coefficients are about independent of frequency. This is ordinarily true above a few hundred cycles. The total far-end coefficient (curve *A*) is about independent of frequency in the important carrier frequency range. The direct far-end coefficient of curve *C'* decreases considerably with frequency for reasons discussed in Appendix A. Since transpositions are ordinarily designed for the condition of a number of wires on a line, the total crosstalk coefficient is the one usually used in practice.

Curves *C* of Fig. 11 also indicate that the direct near-end coefficient is much larger than the direct far-end coefficient. This is usually true and, as discussed in detail in Appendix A, the explanation is that the crosstalk currents caused by the electric and magnetic fields add almost directly in the case of direct near-end crosstalk but tend to cancel in the case of direct far-end crosstalk. As discussed in the appendix, the indirect (vector difference of curves *A* and *C*) crosstalk in a very short length is due almost entirely to the electric field of the tertiary circuits and is the same for both near-end and far-end crosstalk. In Fig. 11, the total near-end coefficient (curve *A*) is increased by the indirect crosstalk since curve *C* is lower than curve *A*. The reverse is usually true, however. In the case of far-end crosstalk the total coefficient is usually increased by the indirect crosstalk.

Crosstalk coefficients are vector quantities and may be measured in magnitude and phase. If it is desired to compute the crosstalk between two long pairs of wires which do not change their pin positions, it is only necessary to know the magnitude of the crosstalk coefficient, since the problem is to determine the ratio of the crosstalk for many elementary lengths to the crosstalk for one such length. However, if it is desired to know the crosstalk between long circuits which do change their pin positions, several crosstalk coefficients must be known, one for each combination of pin positions. In order to determine the total crosstalk for several segments of a line involving different pin

positions, it is necessary to know both the phase and magnitude of the crosstalk coefficients. For practical purposes, however, the coefficients may, in most cases, be regarded as algebraic quantities having sign but not angle.

The direct component of the total crosstalk coefficient may be readily computed as discussed in Appendix A. If more than a very few wires are involved, an exact calculation of the indirect component is impracticable but a fair approximation may be obtained by the method discussed in Appendix A. This method is used when a wire configuration is under consideration but is not available for measurement.

As pointed out in 1907 by Dr. G. A. Campbell, an accurate calculation of the total crosstalk coefficient would involve determination of the "direct capacitances" between wires of the test pairs. Since these capacitances are functions of the distances between all combinations of wires on the lead and between wires and ground, their calculation is usually impracticable. In the past, the crosstalk coefficients were computed by a method proposed by Dr. Campbell which involved measurement of the direct capacitances.[3]

The part of the coefficient due to the electric field was computed from the "direct capacitance unbalance." The part due to the magnetic field was computed as discussed in Appendix A. When loaded open-wire circuits were in vogue it was necessary to be able to separate the electric and magnetic components of the coefficients. After loading was abandoned this separation was unnecessary and it was found more convenient to measure the total coefficients than to measure the direct capacitances or differences between pairs of these capacitances.

As previously discussed, in designing transpositions it is necessary to compute the interaction type of crosstalk indicated by Fig. 2C, and it is, therefore, necessary to have some coupling factor for use in this computation. Such a coupling factor could, theoretically, be determined as indicated schematically by Fig. 12. The interaction crosstalk between two short lengths of line would be measured by transmitting on one pair and receiving on the other pair at the junction of the two short lengths as indicated by the figure.

If there were but a single tertiary circuit such as c of the figure, the crosstalk measured would be that due to the compound crosstalk path $n_{ac}n_{cb}$. In this product, n_{ac} is the near-end crosstalk between a and c in the right-hand short length d and n_{cb} is the near-end crosstalk between c and b in the left-hand short length. Since n_{ac} and n_{cb} when

[3] See papers by Dr. Campbell and Dr. Osborne listed under "Bibliography."

expressed in crosstalk units are current ratios times a million, their product $n_{ac}n_{cb}$ is a current ratio times a million squared. The crosstalk measured would be this current ratio times a million or $n_{ac}n_{cb}10^{-6}$.

For small values of d, n_{ac} and n_{cb} vary directly as the frequency and as the length d. Therefore:

$$n_{ac} = N_{ac}Kd,$$
$$n_{cb} = N_{cb}Kd,$$
$$n_{ac}n_{cb}10^{-6} = N_{ac}N_{cb}K^2d^2 10^{-6},$$

where N_{ac} and N_{cb} are the near-end crosstalk coefficients, K is the frequency in kilocycles and d is expressed in miles. The measured

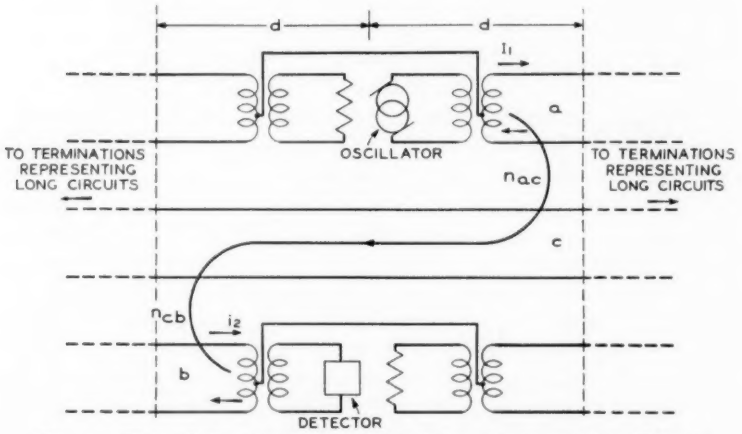


Fig. 12—Theoretical method of measuring interaction crosstalk coefficient.

crosstalk $n_{ac}n_{cb}10^{-6}$ divided by K^2d^2 gives the quantity $N_{ac}N_{cb}10^{-6}$ which may be designated as I_{ab} and called *the interaction crosstalk coefficient.* Values of I_{ab} determined from crosstalk measurements on multi-wire lines would include the effect of numerous tertiary circuits instead of that of a single tertiary circuit as indicated by Fig. 12.

While the interaction crosstalk coefficient I_{ab} could theoretically be measured as outlined above, it is simpler to deduce an approximate value from the measured value of the far-end crosstalk coefficient F_{ab}. The indirect component of F_{ab} is due to the tertiary circuits and must, therefore, be related to I_{ab} which is also due to these circuits. As discussed in detail in Appendix A:

$$I_{ab} = -\frac{2\gamma_c F_{ab}}{K} \text{ approximately.}$$

In this expression K is the frequency in kilocycles and $\gamma_c = \alpha_c + j\beta_c$ is the propagation constant of the tertiary circuit c. On a multi-wire line there would be numerous tertiary circuits with various values of γ. With practicable wire sizes the attenuation constants indicated by α are small compared with the phase change constants indicated by β. Measurements of crosstalk indicate that the values of β are all in the neighborhood of the value given by the expression $\pi K/90$. This corresponds to a speed of propagation of 180,000 miles per second which is about the average for the present carrier frequency range. Neglecting the attenuation constants:

$$\gamma_c = j\beta = j\frac{\pi K}{90},$$

$$I_{ab} = -j\frac{2\pi F_{ab}}{90}.$$

This relation is much used in transposition design. As noted above, the indirect component of F_{ab} should, strictly speaking, be used to obtain I_{ab}. In most cases, however, the total value of F_{ab} may be used since this total is determined largely by the indirect component.

Type Unbalance

A conception important in transposition design is that of "type unbalance." This conception will now be explained and the general method of computation will be discussed.

As we have seen, any two open-wire circuits tend to crosstalk into each other due to coupling between them. By transposing the circuits, the coupling in any short length of line is nearly balanced in another short length by a second coupling of about the same size but about opposite in phase. This balancing is never perfect and there is always a residual unbalanced coupling due to (1) attenuation and change in phase of the disturbing transmission current and resulting crosstalk currents as they are propagated along the circuits and (2) irregularities in the spacing of the transpositions and irregularities in the spacings between the various wires. The term "type unbalance" has been chosen to indicate the residual unbalance caused by propagation effects. It is expressed as an "equivalent untransposed length," that is, the type unbalance times the crosstalk per mile gives the residual crosstalk due to propagation effects assuming no constructional irregularities.

The method of computing the type unbalance for near-end crosstalk will now be discussed. The part of the near-end crosstalk due to interaction between all the different thin slices of line may be ignored

since, as discussed in connection with Figs. 2C and 2D, the interaction crosstalk involves the product of a near-end crosstalk path and a far-end crosstalk path. This product is small since the coupling through the far-end path is inherently small. Therefore, the interaction crosstalk coefficient is much smaller for near-end crosstalk than for far-end crosstalk, while for the transverse crosstalk coefficients the reverse is true.

As was indicated by the discussion of Fig. 8B, the transverse near-end crosstalk between two long circuits may be computed by dividing the parallel into short segments, each having the same transverse crosstalk coupling. The coupling between circuit terminals for any segment will be different from that at the segment terminals due to propagation effects as explained in connection with Fig. 8B. Therefore, the coupling at the circuit terminal for each segment must be determined and, finally, the sum of the coupling values for all the segments.

The simplest case is that of two non-transposed circuits. The problem is indicated by Fig. 13 which is like Fig. 8B except that more segments are shown.

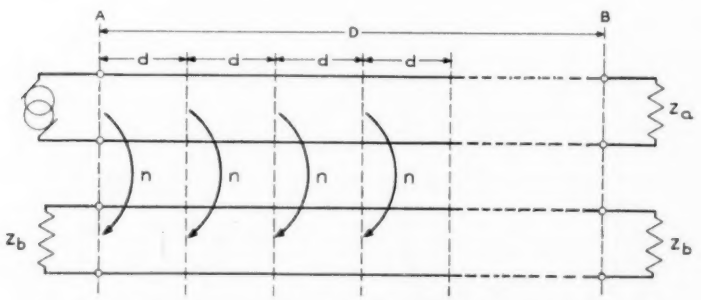


Fig. 13—Method of computing near-end crosstalk between untransposed circuits in length D.

The near-end crosstalk coupling n at point A due to the first segment is NKd, where N is the crosstalk coefficient and K is the frequency in kilocycles. The crosstalk current from the second segment relative to that from the first segment is attenuated by the factor $\epsilon^{-(\alpha_1+\alpha_2)d}$, and also retarded in phase by the angle $\epsilon^{-j(\beta_1+\beta_2)d}$. In other words, the crosstalk current from the second segment is equal to the crosstalk current from the first segment times the factor $\epsilon^{-(\gamma_1+\gamma_2)d}$, where γ_1 and γ_2 are the propagation constants for the two circuits and γ equals $\alpha + j\beta$. Letting γ be the average propagation constant, the coupling

at point A for the second segment is equal to that for the first segment times $\epsilon^{-2\gamma d}$ or $NKd\epsilon^{-2\gamma d}$. The coupling at point A for the third segment is $NKd\epsilon^{-4\gamma d}$. The sum of the crosstalk couplings at point A at all the segments is, therefore:

$$NKd(1 + \epsilon^{-2\gamma d} + \epsilon^{-4\gamma d} + \epsilon^{-6\gamma d} + \text{etc.}).$$

This expression may be summed up for the number of segments corresponding to the total length D. It is simpler, however, to let d be an infinitesimal length and to integrate over the length D, i.e., from point A to point B of Fig. 13. This gives for the total near-end crosstalk for non-transposed circuits:

$$NK\frac{1 - \epsilon^{-2\gamma D}}{2\gamma}.$$

In the special case when D is only the usual short segment between transposition poles, the above expression is practically equal to NKD.

The near-end crosstalk between circuits having transposition poles spaced a considerable distance D apart may now be computed. Figure 14 shows a length $2D$ in a parallel between two long circuits, there being a transposition in one circuit at the center of $2D$. The near-end crosstalk for the length AB is given by the above expression. The near-end crosstalk at point A for the length BC will be the same expression multiplied by the propagation factor $\epsilon^{-2\gamma D}$ and reversed in sign due to the effect of the transposition. The near-end crosstalk at point A for the length $2D$ will, therefore, be the sum of the values for lengths AB and BC. This sum is:

$$NK\frac{1 - \epsilon^{-2\gamma D}}{2\gamma}(1 - \epsilon^{-2\gamma D}).$$

This quantity divided by NK is the type unbalance for the length $2D$ of Fig. 14. If D is only the length of a short segment the above expression is about equal to $NKD(2\gamma D)$.

Similarly the near-end crosstalk at point A for a length $3D$ will be:

$$NK\frac{1 - \epsilon^{-2\gamma D}}{2\gamma}(1 - \epsilon^{-2\gamma D} \mp \epsilon^{-4\gamma D})$$

and the type unbalance is this quantity divided by NK. For a length $4D$ the quantity in the parentheses becomes $(1 - \epsilon^{-2\gamma D} \mp \epsilon^{-4\gamma D} \mp \epsilon^{-6\gamma D})$, etc. The sign of each term in the parentheses is determined by the arrangement of "relative" transpositions, i.e., those at points where only one of the two circuits is transposed. Each term corre-

sponds to a length D. The transposition at the start of the second length (at point B of Fig. 14) reverses the sign of the term for the second length and also the signs for the following lengths until another transposition is reached which makes the next sign plus, etc.

A practical open-wire line is divided into a series of "transposition sections" of eight miles or less. In each section the crosstalk between any two circuits is approximately balanced out by means of transpositions. A main purpose of this division into sections is to provide suitable points for circuits to drop off the line. A circuit on the line for a part of a section may have more crosstalk to a through circuit than if the parallel extended for the whole section since coupling in

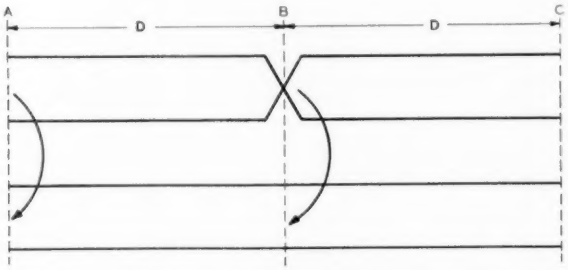


Fig. 14—Near-end crosstalk in length $2D$ between circuits a and b with circuit a transposed in the middle.

the last part of the section may tend to subtract from the coupling in the first part. The ends of sections are, therefore, the most suitable points for circuits to leave or enter the line. Ideally, the sections in a line should all be alike as regards length and transposition arrangements since this makes it practicable to so design the transpositions that residual crosstalk in one section tends to cancel that in another section. Practically, the sections vary in length and, therefore, in the transposition arrangements because the ends of some of the sections must fall at particular "points of discontinuity" determined by branching circuits and by requirements for balance against induction from power circuits.

In designing the transposition sections, type unbalances are computed for the section lengths of eight miles or less. For such lengths, the general method of computing type unbalances may be simplified. The general method involves the vector propagation constant γ. For a length as short as a single transposition section, attenuation can, ordinarily, be neglected. Therefore, in the type unbalance formulas γ can be replaced by $j\beta$ which greatly simplifies the computations.

Since attenuation can be neglected, the type unbalance for a transposition section depends only on the line angle βD. Since β increases practically directly with frequency, a plot of type unbalance against βD indicates the variation of type unbalance with frequency for a fixed length or the variation with length for a fixed frequency. It is convenient to plot the product of type unbalance and frequency (in kilocycles) since this product multiplied by the crosstalk coefficient gives the crosstalk. Two such plots for near-end type unbalance times frequency are shown on Fig. 15A. The plot marked *P* is for the condition of two circuits non-transposed or transposed alike. The plot marked *O* is for the same arrangement except for one relative transposition at the midpoint of the parallel.[4] The figure has a frequency scale corresponding to a length of eight miles as well as the general βD scale in degrees.

It will be seen that, for the case of no relative transpositions, the crosstalk varies directly with the frequency for only a short distance at the start of the curve. The effect of one relative transposition is to greatly reduce the crosstalk for small values of βL. For larger values the crosstalk is increased. It may be noted that the minimum values shown on the curves are somewhat in error since attenuation was neglected.

The minimum values in the *P* curve are due to "natural transpositions" in the non-transposed circuits. When the line angle is 180 degrees the crosstalk at the near-end of the disturbed circuit due to the second half of the line is just 180 degrees out of phase with the crosstalk due to the first half. This reversal in phase is due to the phase change accompanying the propagation of current to the mid-point and back. The total crosstalk due to both halves of the line lengths is the same as if the crosstalk coupling in the second half were translated to the near-end and the parallel without phase change but one circuit was transposed at the mid-point. When the line angle is 360 degrees the "natural transpositions" are at the quarter points, etc.

The near-end crosstalk between any two circuits in a transposition section may be estimated by multiplying the crosstalk coefficient by values of type unbalance times frequency similar to those of Fig. 15A. The total crosstalk in a succession of similar transposition sections is calculated at any particular frequency by working out a factor similar to the type unbalance in order to obtain the relation between the crosstalk in many transposition sections and that in one section. In calculating this factor, attenuation cannot be neglected since long lengths of line are involved.

[4] Two circuits are relatively transposed by one transposition at a given point in the line. Transpositions in both circuits leave them relatively untransposed.

The method of computing type unbalances for far-end crosstalk will now be explained. As in the case of near-end crosstalk, the type unbalance is defined by expressing the far-end crosstalk between two long circuits as the product of the crosstalk coefficient, the frequency in kilocycles and the type unbalance.

Figure 15B indicates the periodic variation with frequency of the far-end crosstalk when type unbalance is controlling.

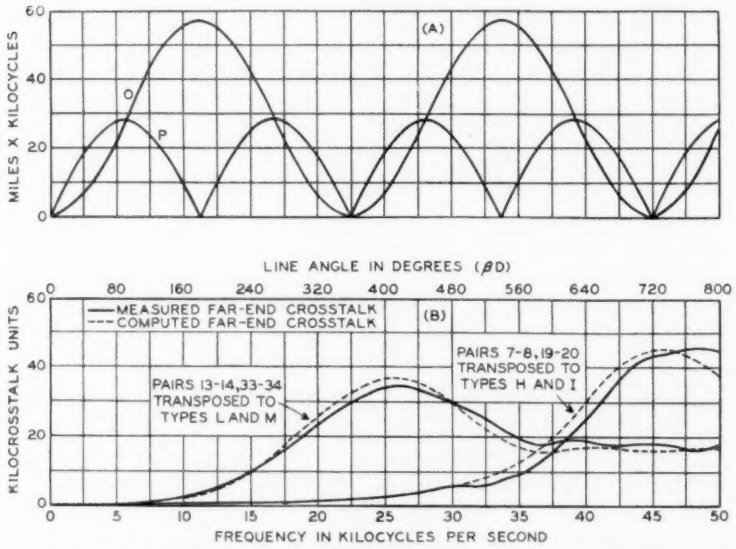


Fig. 15—Type unbalance and crosstalk vs. frequency and line angle in degrees. For Part (*B*), see Fig. 27*A* for wire configuration and Fig. 28 for transposition types.

In the case of near-end crosstalk, the method of computing the type unbalance neglected interaction crosstalk since, ordinarily, the transpositions needed to control transverse crosstalk make the interaction effect negligible. In the case of far-end crosstalk, the most important type of interaction crosstalk is included in calculations of type unbalances but another type of interaction crosstalk and the direct transverse crosstalk are neglected. The transpositions needed to properly suppress the important type of interaction crosstalk and the indirect transverse crosstalk ordinarily make the neglected types of crosstalk very small and the application of a more precise method of computing type unbalances for far-end crosstalk is not justified in practice.

The far-end type unbalance for a non-transposed part of a long parallel between two circuits will be computed first. Such a part of a parallel is indicated by length D of Fig. 16. For purposes of computation this length is divided up into a number of short segments each of length d. Considering the far-end crosstalk for two such segments at the start of the length D it will be seen from the discussion of crosstalk coefficients that transverse crosstalk in the length $2d$ will be

$$2FKd = 2(F_d + F_i)Kd.$$

In the above expression F is the far-end crosstalk coefficient, F_d being that part due to direct crosstalk and F_i that part due to indirect crosstalk.

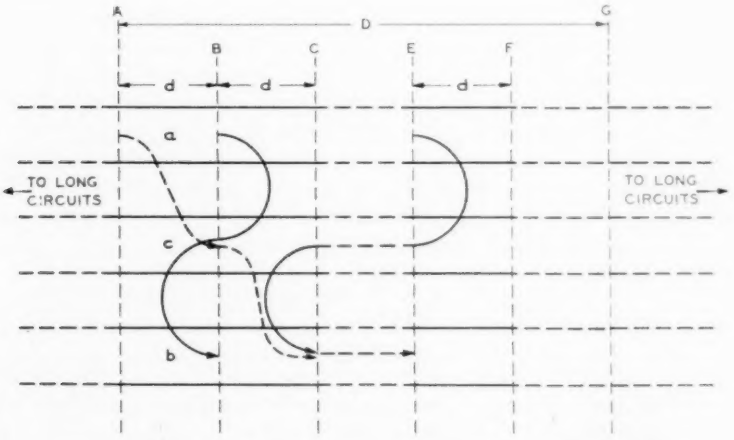


Fig. 16—Far-end crosstalk between untransposed circuits in length D.

The above expression relates to the output-to-output crosstalk. The input-to-output crosstalk is obtained by multiplying by the propagation factor $\epsilon^{-2\gamma d}$ to allow for propagation from A to C. This correction is usually made only when it is desired to obtain the input-to-output crosstalk between complete circuits and it is usually satisfactory to correct by using the attenuation factor and ignoring change in phase.

The total transverse output-to-output crosstalk in the length D is:

$$(F_d + F_i)KD.$$

This is about equal to F_iKD since F_d is ordinarily small compared to F_i.

Figure 16 indicates with a solid line the important type of interaction crosstalk between the first two segments by way of a representative tertiary circuit c. As discussed in the section on crosstalk coefficients and in Appendix A, the far-end crosstalk (output-to-output) of this interaction type will be

$$N_{ac}N_{cb}K^2d^2 10^{-6} = -2\gamma F_i K d^2 \text{ approximately.}$$

The interaction crosstalk as well as the transverse crosstalk is about proportional to the indirect coefficient F_i.

Each segment of the disturbing circuit will have a similar interaction crosstalk coupling with each preceding segment of the disturbed circuit. The interaction crosstalk between segment EF and segment BC is indicated on Fig. 16. The expression for this differs from the above expression in that the additional propagation distance from E to C and back must be allowed for. To get the total output-to-output far-end crosstalk it is necessary to sum up all these interaction crosstalk couplings between segments and to this sum add the total transverse crosstalk in length D.

This clumsy summation process may be avoided by letting d be an infinitesimal length and integrating between points A and G. This results in the following approximate expression for the output-to-output far-end crosstalk in the length D.

$$F_d K D + F_i K D + F_i K \left[\frac{1 - \epsilon^{-2\gamma D}}{2\gamma} - D \right].$$

This assumes the same propagation constant for the disturbing, disturbed and tertiary circuits. This approximation is justified for short lengths of, say, 10 miles or less.

The last term represents the interaction crosstalk and this term is negligible for small values of D. For larger values of D interaction crosstalk must be considered and it is convenient to rewrite the expression as follows:

$$F_d K D + F_i K \frac{1 - \epsilon^{-2\gamma D}}{2\gamma}.$$

The first term representing the direct crosstalk is negligible for values of D corresponding to a line angle of 90 degrees or less since F_d is ordinarily small compared with F_i and D is not large compared with $(1 - \epsilon^{-2\gamma D}/2\gamma)$. Therefore, direct crosstalk ordinarily may be neglected in computing far-end type unbalance. Another reason for neglecting direct crosstalk is that it is readily cancelled by a few relative transpositions while the remaining far-end crosstalk depends

upon the transpositions in a complicated way, because the various interaction crosstalk couplings involve a variety of propagation distances and, therefore, have a variety of phase angles. If both circuits are transposed frequently but alike the direct crosstalk is not affected by the transpositions but it is ordinarily small compared with the indirect transverse crosstalk.

Figure 16 indicates by a dashed line another type of interaction crosstalk involving the product of two far-end crosstalk couplings. This effect can be neglected with practical arrangement of transpositions but may be important in the case of circuits having few transpositions or none at all.

In computing type unbalance the far-end crosstalk in an untransposed segment of line of length D may, therefore, be written as:

$$F_i K \frac{1 - \epsilon^{-2\gamma D}}{2\gamma} = FK \frac{1 - \epsilon^{-2\gamma D}}{2\gamma} \text{ approx.}$$

Since the magnitude of F_i is ordinarily about equal to that of F, the measured coefficient, it is usually satisfactory to use the latter value.

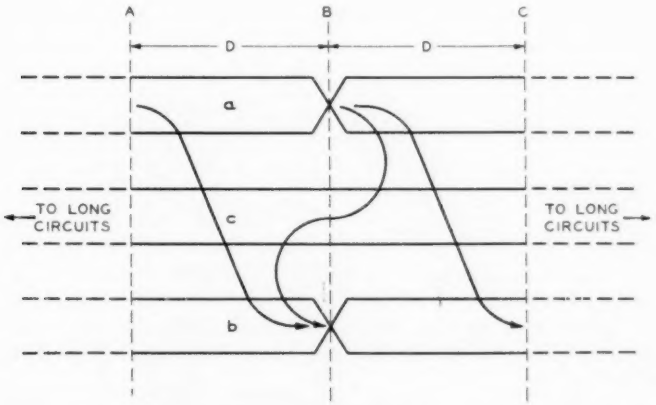


Fig. 17—Far-end crosstalk in length $2D$ between circuits a and b with each circuit transposed at the middle.

Having derived the above expression it is now possible to derive the far-end type unbalance for two transposed circuits. Figure 17 indicates a parallel between two long circuits. The type unbalance will be computed for a length $2D$ in which both circuits are transposed at the center. In the length $2D$ three far-end crosstalk paths must be considered, that is, the far-end crosstalk in length AB, that in length

BC and the important type of interaction crosstalk between length BC and length AB. The output-to-output crosstalk values only will be written for all these paths or, in other words, the effect of the propagation distance AC will not be considered in the expressions.

The far-end crosstalk for either length AB or BC is given by the above expression. Since *both* circuits are transposed at point B the far-end crosstalk values in the two lengths will add directly and their sum will be

$$2FK\frac{1-\epsilon^{-2\gamma D}}{2\gamma}.$$

Transmission from A to C through the crosstalk path in length AB is reversed in sign due to the transposition in circuit b at B. The output current of circuit a is also reversed in sign. In general, the output-to-output current ratio may or may not be reversed in sign depending on the transposition arrangement. It is convenient, however, to consider the first path as a reference and assign a plus sign to the crosstalk. Other paths are then assigned the proper relative phase angles.

As discussed in connection with Fig. 16, if the length D is very short the interaction crosstalk between the two segments may be written:

$$-2\gamma F_iKD^2 = -2\gamma FKD^2 \text{ approx.}$$

In practice the length D may be too long for this approximate expression in which case it is necessary to substitute for D in the above expression the value derived in connection with the discussion of the near-end crosstalk in a length D. In other words, D of the above expression should be replaced by

$$\frac{1-\epsilon^{-2\gamma D}}{2\gamma}.$$

With this substitution the interaction crosstalk between the two lengths becomes

$$-FK\frac{(1-\epsilon^{-2\gamma D})^2}{2\gamma}.$$

Transmission from A to C through this crosstalk path involves two transpositions and therefore the sign of the above expression is not reversed. Relative to the reference path through the crosstalk in length AB the sign should be reversed, however, and become plus. The total crosstalk in the length $2D$ is, therefore,

$$2FK\frac{1-\epsilon^{-2\gamma D}}{2\gamma}+FK\frac{(1-\epsilon^{-2\gamma D})^2}{2\gamma}=FK\frac{3-4\epsilon^{-2\gamma D}+\epsilon^{-4\gamma D}}{2\gamma}.$$

The latter expression divided by F is the frequency times the far-end type unbalance for the length $2D$. If one of the circuits were transposed at point B the crosstalk in length AB would be cancelled by that in length BC. The sign of the interaction crosstalk between the two lengths would be reversed and the expression would become

$$-FK\frac{(1-\epsilon^{-2\gamma D})^2}{2\gamma}.$$

If neither circuit were transposed at B, the far-end crosstalk would be that for a non-transposed length of $2D$ or:

$$FK\frac{1-\epsilon^{-4\gamma D}}{2\gamma}.$$

The frequency times the type unbalance values for the cases of one transposition and no transpositions are the same (in magnitude) as those derived for near-end crosstalk which were plotted (neglecting attenuation) as curves O and P on Fig. 15A.

If both circuits are transposed at B the near-end type unbalance remains the same as if there were no transpositions. The far-end type unbalance is radically altered, however. This is evident if the above equation is compared with that for the case of both circuits transposed.

This process of computing type unbalances may be extended from two equal lengths to any number of equal lengths. It is necessary to consider the interaction crosstalk between each length of the disturbing circuit and each preceding length of the disturbed circuit. The relative propagation distances through the various interaction crosstalk couplings must be taken account of.

Computations of far-end type unbalances are greatly simplified by assuming the same propagation constants for the disturbing, disturbed and tertiary circuits and by neglecting attenuation within a transposition section as in the case of near-end crosstalk. Since the tertiary circuit may be composed of any combination of wires on the line or of these wires and ground return, the propagation constant for a tertiary circuit may be somewhat different from that for the disturbing and disturbed circuits. This is particularly true of earth-return circuits, but these are of little practical importance due to their relatively high attenuation. All circuits not involving the earth have somewhere near the same speed of propagation but the tertiary circuits may differ greatly in attenuation constants.

For practical reasons a fair balance against crosstalk must be obtained in each transposition section (eight miles or less) and, as in the case of near-end crosstalk, type unbalances are calculated for the

transposition arrangements which may exist in a single transposition section. Since the attenuation in a transposition section is not great, these calculations need not take into account differences in the attenuation constants of the various tertiary circuits. A long line has a series of transposition sections of various types and the total far-end crosstalk for any two circuits is a summation of the crosstalk values obtained from the type unbalances for the various sections plus interaction crosstalk between the various combinations of sections. With practical methods of transposition design, the transposition arrangements are so chosen that the interaction crosstalk between two sections is usually small compared with the far-end crosstalk in one section. A long line for the most part consists of a succession of similar sections with occasional sections of other types. Interaction crosstalk between dissimilar sections does not ordinarily contribute appreciably to the total far-end crosstalk. For the important case of a succession of similar sections interaction crosstalk between sections must be carefully considered since it may build up systematically and the total may be large compared with the summation for the far-end crosstalk values for the individual sections.

Serious interaction crosstalk between similar sections is guarded against by computing factors relating the far-end type unbalance in one section to that in various numbers of successive sections with various transposition arrangements at the junctions of sections. The factors actually computed are somewhat in error since they involve long distances and assume the same attenuation constants for all circuits. The errors are not sufficient, however, to prevent the factors from being a proper guide in avoiding systematic building up of interaction crosstalk between sections.

The above discussion assumes that the tertiary circuits are indefinitely extended or terminated to simulate their characteristic impedance. The tertiary circuits may not be terminated at the ends of a line since many of them are not used for transmission of speech or signals. Complete reflections of the crosstalk current in the tertiary circuits will, therefore, occur at their ends and these reflections somewhat modify the crosstalk currents in other circuits. This effect is important in a very short line since the reflected wave is again reflected at the distant end and at particular frequencies large changes in the tertiary crosstalk currents may occur due to multiple reflections. In a long line such multiple reflections are damped out and, in general, tertiary circuit reflection effects are not important.

If all the pairs on a line are transposed for the same maximum useful frequency, the transposed pairs will usually be relatively

unimportant as tertiary circuits, that is, two pairs having small crosstalk between them usually contribute but little to the crosstalk between one of these pairs and any third pair. In some cases, however, this effect is important. On some lines certain pairs may be transposed for carrier operation and other circuits on the line for voice frequencies only. A combination of the two kinds of circuits may have large crosstalk between them at carrier frequencies and may contribute appreciably to the carrier frequency crosstalk between the pair transposed for carrier operation and some other pair also so transposed.

Far-end type unbalances which take account of transpositions in a tertiary circuit must, therefore, be calculated. This can be done by following the same general method discussed in connection with Fig. 17. From the discussion of coefficients it follows that the far-end coefficient for use in computing such a type unbalance will be:

$$\frac{-N_{ac}N_{cb}K}{2\gamma}\,10^{-6},$$

where N_{ac} and N_{cb} are the near-end crosstalk coefficients for the combination of disturbing circuit and tertiary circuit and the combination of tertiary circuit and disturbed circuit. Since these circuit combinations involve recognized transmission circuits, their near-end coefficients will be available since they must be measured or computed in order to compute the near-end crosstalk.

If a parallel between two circuits is divided into a large number of segments by transposition poles there is a wide variety of transposition arrangements which may be installed at these poles. It is, therefore, a complicated problem to devise charts and tables in reasonable numbers which will cover all the possible type unbalance values for the various transposition arrangements over a wide range of frequencies. This is particularly true in the case of far-end type unbalances since the type unbalance is altered by transposing both circuits at the same points and it is necessary to work out a type unbalance for each combination of transposition arrangements which may be used in two circuits. In the case of near-end crosstalk a number of different transposition arrangements will have the same type unbalance since only the relative transpositions need be considered.

The circuit capacity of a line may be increased by the use of phantom circuits (generally when carrier-frequency systems are not involved) which must, of course, be transposed to avoid noise and crosstalk. The crosstalk between phantom circuits may be calculated in a manner similar to that for pairs. The calculation of crosstalk between side

circuits of the phantoms or between a side circuit and a phantom circuit is complicated by the fact that the phantom transpositions cause the side circuits to change pin positions. Near-end and far-end type unbalances have been computed, however, which take account of this "pin shift" effect of the phantom circuits. In general, the use of phantom circuits seriously limits the crosstalk reduction which may be obtained by transpositions. Phantom circuits are often uneconomic since they seriously restrict the number of carrier frequency channels which may be operated over a given pole line.

As indicated by Fig. 15A the values of type unbalance times frequency have marked maximum and minimum values when they are plotted against frequency or length. The maximum values are usually reduced by increasing the number of transpositions in a given length. When there are a number of circuits on the line it is usually necessary that the propagation of current between successive transposition poles does not change the phase by more than about five degrees. Since the phase change is about two degrees per mile per kilocycle the maximum transposition interval in miles is about $2.5/F$ where F is the frequency in kilocycles. This means .25 mile or 1300 feet at 10 kilocycles and .06 mile or 300 feet at 40 kilocycles.

It does not follow, however, that the least maximum value of type unbalance for a range of frequencies is obtained by using the greatest number of transpositions for a given number of transposition poles. This is illustrated by Fig. 15A which shows that the least maximum value is obtained with no transpositions rather than with one transposition. The total crosstalk current at a terminal is composed of numerous elements of various magnitudes and phase relations. The vector sum of these elements tends to be small at particular frequencies with no transpositions at all and it is important to preserve this tendency as much as possible when choosing an arrangement of transpositions. The vector sum of the elements can never be made zero since this would require that the circuits have no attenuation and infinite speed of propagation. This sum and, therefore, the type unbalances can be made very small, however, by choosing a suitable transposition arrangement and making the interval between transposition poles very small. In practice, the values of type unbalance times frequency for adjacent circuits are restricted to values much less than those of Fig. 15A.

Vacuum Tube Electronics at Ultra-high Frequencies *

By F. B. LLEWELLYN

Vacuum tube electronics are analyzed when the time of flight of the electrons is taken into account. The analysis starts with a known current, which in general consists of direct-current value plus a number of alternating-current components. The velocities of the electrons are associated with corresponding current components, and from these velocities the potential differences are computed, so that the final result may be expressed in the form of an impedance.

Applications of the general analysis are made to diodes, triodes with negative grid, and to triodes with positive grid and either negative or positive plate which constitute the Barkhausen type of ultra-high-frequency oscillator. A wave-length range extending from infinity down to only a few centimeters is considered, and it is shown that even in the low-frequency range certain slight modifications should be made in our usual analysis of the negative grid triode.

Oscillation conditions for positive grid triodes are indicated, and a brief discussion of the general assumptions made in the theory is appended.

I. Foreword

THE art of producing, detecting, and modulating ultra-high-frequency electric oscillations has reached the same state of development which was attained in early work on lower frequency oscillations when experiment had outstripped theory. The experimenters were able to produce oscillations by using vacuum tubes, but were not able to explain why. They were able to make improvements by the long and tedious process of cut and try, but did not have the powerful tools of theoretical analysis at their command. In particular, the advantage of the theoretical attack may be illustrated by the rapid advance in technique which followed the theoretical concept of the internal cathode-plate impedance of three-element vacuum tubes. The work of van der Bijl and Nichols showed that for purposes of circuit analysis this path could be replaced by a fictitious generator of voltage, μe_g, having an internal impedance whose magnitude is given by the reciprocal of the slope of the static $V_p - I_p$ characteristic. Development of commercially reliable vacuum tube circuits began forthwith. In a similar, yet less complicated manner, the internal network of two-element tubes may be replaced by an equivalent resistance when relatively low frequencies only are considered.

In these concepts where the vacuum tube is replaced by its equiva-

* Presented in brief summary before U. R. S. I., Washington, D. C., April, 1932. *Proc. I. R. E.*, Vol. 21, No. 11, November, 1933.

lent network impedance, one outstanding feature is exemplified: namely, the separation of the alternating- and direct-current components. The equivalent networks are applicable to the alternating-current fundamental component of the current and differ widely from the direct-current characteristics. A complete realization of the importance of this separation will be of advantage in the later steps where extension of the classical theory to the case of ultra-high-frequency currents is described.

For a short time after the original introduction of the equivalent network of the tube, affairs progressed smoothly. Soon, however, frequencies were increased and a new complication arose. The difficulty was attributable to the interelectrode capacities existing between the various elements of the vacuum tube. The original attempts to take this into account were based on the viewpoint that the tube network should be complete in itself and separate from the external circuit network to which it was attached. Correct results, of course, were obtained by this method but later developments showed the advantage of considering the equivalent network of the complete circuit, including both tube and external impedances in a single network. For instance, by grouping the combination of grid-cathode capacity with whatever external impedance was connected between these two electrodes, a great simplification occurred. This step also has its analogy in the development of ultra-high-frequency relations.

As time went on, higher and higher frequencies were desired, and they were produced by the same kind of vacuum tubes operating in the same kind of circuits, although refinements in circuit and tube design allowed the technique to be improved to the point where oscillations of the order of 70 to 80 megacycles were obtainable with fair efficiency. When the frequency was increased still further, it was found that extension of the same kind of refinements was unavailing in maintaining the efficiency and mode of operation of the higher frequency oscillations at the level which had previously been secured. Ultimately, the three-electrode tube regenerative oscillator ceases to function as a power generator in the neighborhood of 100 megacycles for the more usual types of transmitting tubes. When this point was reached, the external circuit had not yet shrunk up to zero proportions and neither had its losses become sufficiently high to account altogether for the failure of the tube to produce oscillations. From this point on, the old-time cut-and-try methods were employed and marked improvements were secured. In fact, low power tubes have been made which operate at wave-lengths of the order of 50 to 100 centimeters with fair stability, although quite low efficiency.

In the meantime, the production of ultra-high-frequency oscillations had been progressing in a somewhat different direction. The discovery, about 1920, by Barkhausen that oscillations of less than 100 centimeters wave-length could be secured in a tube having a symmetrical structure, when the grid was operated at a fairly high positive potential, while the plate was approximately at the cathode potential, started experiments on what was thought to be an altogether different mode of oscillation. Workers by the score have extended both the experimental technique and the theory of production of this newer type of oscillation. However, one of the results which an analysis of ultra-high-frequency electronics illustrates is that the electron type of oscillator is merely another example of the same kind of oscillation which was produced in the old-time so-called regenerative circuits.

For the purpose of extending the theory of electronics within vacuum tubes to frequencies where the time of transit of the electrons becomes comparable with the oscillation period, it is important at the outset to select an idealized picture which is simple enough to allow exact mathematical relations to be written. At the same time, the picture must be capable of adaptation to practical circuits without undue violence to the mathematics. An example of this kind of adaptation is illustrated by the classical calculation of the amplification factor μ, which was accomplished by consideration of the force of the electrostatic field existing near the cathode in the absence of space charge even though tubes were never operated under this condition. In a like manner, such violations of the ideal must, of necessity, be made in ultra-high-frequency analysis but their practical validity lies in so choosing them that the quantitative error introduced is less than the expected precision of measurement. It becomes, therefore, of the utmost importance to state clearly the transitions which occur between results obtained for the idealized case to which the mathematics is strictly applicable and the practical circuits where the assumptions and approximations are made to conform with operating conditions.

A start has already been made on the problem of developing such a generally valid system of electronics. This was done by Benham [1] who considers a special case comprising two parallel-plane electrodes, one of which is an emitter and the other a collector, when conditions at the emitter are restricted by the assumption that the electrons are emitted with neither initial velocity nor acceleration. This work of Benham's has the utmost importance in a general electronic theory

[1] W. E. Benham, "Theory of the Internal Action of Thermionic Systems at Moderately High Frequencies," Part I, *Phil. Mag.*, p. 641; March (1928); Part II, *Phil. Mag.*, Vol. 11, p. 457; February (1931).

and, in fact, the means of extending his theory exists primarily in the selection of much more general boundary conditions than were assumed by him. It will, therefore, result that some repetition of Benham's work will appear in the following pages. However, in view of the new state of the theory and the importance of accurate foundations for it, this repetition is advantageous rather than otherwise.

With these preliminary remarks in mind, the next step is the selection of the idealized starting point for a mathematical analysis. Exactly as was done by Benham we take two parallel planes of infinite extent, one of which is held at a positive potential V with respect to the other, and between the two electrons are free to move under the influence of the existing fields. The next step in the idealization constitutes the separation of alternating- and direct-current components not only of current and potential, but also of electron velocity, charge density, and electric intensity. With this separation, the restriction that the direct-current component of the electron velocity and acceleration is zero at the negative plane may be made while leaving us free to select much more general boundary conditions for the alternating-current component. It is true that the more general conditions now proposed will not fit the original physical picture where the negative plane consists of a thermionic emitter. Nevertheless the extension is of importance since it allows application to be made to the wide number of physical cases where "virtual cathodes" are formed. One such example is the convergence of electrons toward a plate maintained at cathode potential while a grid operating at a high positive potential with respect to both is interposed between them. In a stricter mathematical sense, the broader boundary conditions come about because of the fact that the general equations containing all components are separable into a system of equations, one for each component, and that the boundary conditions for the different equations of the system are independent of each other.

The concept of an alternating-current velocity component requires a few words of explanation. In the absence of all alternating-current components, electrons leave the cathode with zero velocity and acceleration and move across to the anode with constantly increasing velocity under the well-known classical laws. This velocity constitutes the direct-current velocity component. When the alternating-current components are introduced, there will be a fluctuation in velocity superposed on the direct-current value, and the alternating-current component need not be zero at a virtual cathode. This separation of components will come about naturally in the course of the mathematical analysis which follows, but since the interpretation of the

equations is of paramount importance, a few words of explanation and repetition will be necessary.

II. Fundamental Relations

For the development of the fundamental relations existing between the two parallel planes, we have the classical equations of the electromagnetic theory which may be set down in the following form:

$$\left.\begin{aligned} E &= -\frac{\partial V}{\partial x}, \\ \frac{\partial E}{\partial x} &= 4\pi P, \\ J &= PU + \frac{1}{4\pi}\frac{\partial E}{\partial t}, \end{aligned}\right\} \tag{1}$$

where E is the electric intensity, V the potential, P the charge density, J the total current density consisting of conduction and displacement components, and U is the charge velocity. These equations apply to frequencies such that the time which would be taken by an electromagnetic wave in traveling between the two planes is inappreciable when compared with the period of any alternating-current frequency considered. Ordinarily this limitation will become of importance only at frequencies higher even than those in the centimeter wave-length range where the time of electron transit is of great importance, although the time of passage of an electromagnetic wave is still negligibly small.

An electron situated between the two parallel plates will be acted upon by a force which determines its acceleration. The resulting velocity is a function both of the distance, x, from the cathode and the time, t, so that in terms of partial derivatives, the equation expressing the relation between the force and acceleration is

$$\frac{\partial U}{\partial t} + U\frac{\partial U}{\partial x} = \frac{e}{m}E. \tag{2}$$

From (1) and (2) may readily be obtained

$$\left(\frac{\partial}{\partial t} + U\frac{\partial}{\partial x}\right)^2 U = 4\pi\frac{e}{m}J. \tag{3}$$

In this equation we have a relation between the velocity and the total current density. The advantage of this form of equation for a starting point lies in the fact that the total current density J is not a function

of x. This comes about because of the plane shape and parallel disposition of the electrodes, and the fact that current always flows in closed paths. Thus, while the current between the two planes may be a function of time, it is not a function of x.

The separation of alternating- and direct-current components may now be made. We write

$$J = J_0 + J_1 + J_2 + \cdots \tag{4}$$

with corresponding

$$\left.\begin{aligned} U &= U_0 + U_1 + U_2 + \cdots, \\ V &= V_0 + V_1 + V_2 + \cdots, \end{aligned}\right\} \tag{5}$$

where the quantities with the zero subscript are dependent on x, only, those with subscript 1 are dependent to first order of small quantities upon time, those with subscript 2 are dependent to second order, and so forth. As a result of this separation in accord with the order of dependents upon time, (3) may be split up into a system of equations, the first of which expresses the relation between U_0, J_0, and x and does not involve time. This is the relation governing the direct-current components. The second equation of the system involves the relation between U_1, J_1, x, and time, and contains U_0 which was determined by the first equation. Likewise, the third equation contains U_2, U_1, J_2, x, and t. Since the series given by (4) and (5) are convergent so that, in general, the terms with higher order subscripts are smaller than those with lower subscripts, we may consider that, at least for small values of alternating-current components, the total fundamental frequency component is given by the terms with unity subscript.

The first two equations of the system are as follows:

$$U_0 \frac{\partial}{\partial x}\left(U_0 \frac{\partial U_0}{\partial x} \right) = 4\pi \frac{e}{m} J_0, \tag{6}$$

$$\left(\frac{\partial}{\partial t} + U_0 \frac{\partial}{\partial x} \right)\left(\frac{\partial U_1}{\partial t} + U_0 \frac{\partial U_1}{\partial x} + U_1 \frac{\partial U_0}{\partial x} \right) + U_1 \frac{\partial}{\partial x}\left(U_0 \frac{\partial U_0}{\partial x} \right) = 4\pi \frac{e}{m} J_1. \tag{7}$$

In the solution of (6), the boundary conditions are restricted so that when x is zero, the velocity and acceleration both are zero. These restrictions mean that initial velocities are neglected, and that complete space charge is assumed. Thus the solution for U_0 is

$$U_0 = \alpha x^{2/3}, \tag{8}$$

where

$$\alpha = \left(18\pi \frac{e}{m} J_0 \right)^{1/3}. \tag{9}$$

The solution of (7) is more complicated. We assign a particular value to J_1, namely, $J_1 = A \sin pt$ and find the corresponding value of U_1. To do this, it is convenient to change the variable x to a new variable ξ, which will be called the transit angle. This new variable is equal to the product of the angular frequency p and the time τ which it would take an electron moving with velocity U_0 to reach the point x and is given as follows:

$$\xi = p\tau = \frac{3p}{\alpha} x^{1/3}. \tag{10}$$

Upon changing the dependent variable from U_1 to ω, where $U_1 = \omega/\xi$, we find from (7)

$$\left(\frac{\partial}{\partial t} + p \frac{\partial}{\partial \xi} \right)^2 \omega = \xi\beta \sin pt, \tag{11}$$

where

$$\beta = 4\pi \frac{e}{m} A.$$

This has the solution

$$U_1 = -\frac{\beta}{p^2} \left[\sin pt + \frac{1}{\xi} \cos pt + F_1(\xi - pt) + \frac{1}{\xi} F_2(\xi - pt) \right]. \tag{12}$$

This equation contains two arbitrary functions of $(\xi - pt)$ which must be evaluated by the boundary conditions selected for U_1. Thus the boundary conditions for the alternating-current component make their first appearance.

From the form of (7) which is linear in U_1, it is evident that U_1 must be a sinusoidal function of time having an angular frequency p in order to correspond with the form of J_1. It follows, then, that the most general form which can be assumed for the steady state functions F_1 and F_2 is as follows:

$$\left.\begin{aligned} F_1(\xi - pt) &= a \sin (\xi - pt) + b \cos (\xi - pt) \\ F_2(\xi - pt) &= c \sin (\xi - pt) + d \cos (\xi - pt) \end{aligned}\right\}. \tag{13}$$

Now for the boundary conditions. As pointed out, there is no mathematical necessity for the boundary conditions imposed upon U_1 to correspond with those which were imposed upon U_0. At an actual cathode consisting of an electron emitting surface it would be appropriate to assume that the initial velocities are in no way dependent upon the current, but we shall have to deal not only with actual

cathodes, but also with virtual [2] cathodes where the assumption of zero alternating-current velocity and acceleration is unwarranted. Such a virtual cathode might occur, for instance, between a grid operated at a positive direct-current potential and a plate nearly at cathode potential. If enough electrons came through the mesh of the grid to depress the potential until it became practically zero at some point in the space between grid and plate, the direct-current boundary conditions of zero velocity and acceleration of electrons would be fulfilled at that point. The general equations for the alternating current will therefore apply when the origin is taken at the point of direct-current potential minimum which forms the virtual cathode, and when all of the electrons which are emitted by the actual cathode pass by the virtual cathode and reach the plate. In the event that some of the electrons are turned back at the virtual cathode and move again toward the grid, as indeed they all do when the plate is at a negative potential, a change in the form of the general equation is necessary, and will be described in the sections dealing particularly with positive grid triodes. This change, however, affects merely the form of the equations and not the physical arguments underlying the selection of boundary conditions, which are the same whether all the electrons reach the plate or whether some or all of them turn back toward the grid.

If the alternating-current velocity is determined by small variations in grid potential, let us say, it is evident that no additional assumptions save the requirement that the velocity must not become infinite may be made concerning its value at the virtual cathode. Consequently, a quite general set of boundary conditions will suffice to determine the quantities, a, b, c, d, which appear in (13) and thus completely determine U_1.

Since there are two arbitrary functions in (12), two boundary conditions will be needed. Further inspection shows that the stipulation that the alternating-current velocity be finite at the origin is sufficient to furnish one of these boundary conditions. For the other, a knowledge of the value of the alternating-current velocity at any point between the two reference planes is sufficient. Thus, if at a particular value of ξ, say ξ_1, we know that U_1 is equal to $M \sin pt + N \cos pt$, we have enough information to calculate its value at all other points between the two planes. For example, the two reference planes might be the grid and plate of a positive grid triode. In this event, the alternating-current velocity at the grid could be calculated at the grid plane by means of conditions between there and the cathode.

[2] E. W. B. Gill, "A Space-Charge Effect," *Phil. Mag.*, Vol. 49, p. 993 (1925).

In mathematical form the two boundary conditions may be set forth as follows:

when,

$$\xi = 0, \qquad U_1 \text{ must be finite,} \tag{14}$$

$$\xi = \xi_1, \qquad U_1 = M \sin pt + N \cos pt. \tag{15}$$

From (12) and (13) these result in the values:

$$c = 0, \qquad d = -2,$$

$$a = \frac{p^2}{\beta}(M \cos \xi_1 - N \sin \xi_1) + \cos \xi_1 - \frac{2}{\xi^1} \sin \xi_1, \tag{16}$$

$$b = \frac{2}{\xi_1}(1 - \cos \xi_1) - \sin \xi_1 - \frac{p^2}{\beta}(M \sin \xi_1 + N \cos \xi_1). \tag{17}$$

Thence from (12) we have for the alternating-current velocity, in general,

$$\begin{aligned} U_1 = {} & (M + iN)(\cos \xi_1 + i \sin \xi_1)(\cos \xi - i \sin \xi) \\ & + \frac{\beta}{p^2}\left[\left\{\left(\cos \xi_1 - \frac{2}{\xi_1} \sin \xi_1\right) - i\left(\frac{2}{\xi_1} - \frac{2}{\xi_1} \cos \xi_1 - \sin \xi_1\right)\right\}\right. \\ & \left.(\cos \xi - i \sin \xi) - \left(1 - \frac{2}{\xi} \sin \xi\right) - i\frac{2}{\xi}(1 - \cos \xi)\right], \end{aligned} \tag{18}$$

where, in accord with engineering practice, complex notation is employed, so that sin pt has been replaced by e^{ipt} and cos pt has been replaced by ie^{ipt}, where $i = \sqrt{-1}$.

The first step in the derivation of fundamental relations has now been achieved. The alternating-current velocity at any point between the two planes has been expressed in terms of the alternating-current velocity, $M + iN$, existing at a definite value of x, say x_1, corresponding to the transit angle ξ_1.

The next step is a determination of the potentials corresponding to the velocities U_0 and U_1, respectively. Thus from (1) and (2)

$$-\frac{e}{m}\frac{\partial V}{\partial x} = \frac{\partial U}{\partial t} + U\frac{\partial U}{\partial x} \tag{19}$$

and then with the separation of components as given by (5)

$$-\frac{e}{m}\frac{\partial V_0}{\partial x} = U_0 \frac{\partial U_0}{\partial x}, \tag{20}$$

$$-\frac{e}{m}\frac{\partial V_1}{\partial x} = \frac{\partial U_1}{\partial t} + \frac{\partial}{\partial x}(U_0 U_1). \tag{21}$$

The solution of (20) is

$$V_0 = -\frac{m}{2e} U_0^2 = -\frac{m}{2e} \alpha^2 x^{4/3}, \tag{22}$$

which is the well-known classical relation between the potential, the current, and the position between two parallel planes where complete space charge exists. The complete space-charge condition is postulated by the boundary conditions selected for U_0 and the implications involved are discussed by I. Langmuir and Karl T. Compton.[3]

The alternating-current component of the potential is obtained by integration of (21) as follows:

$$-\frac{e}{m} V_1 = \frac{\partial}{\partial t} \int U_1 \partial x + U_0 U_1 + f(t), \tag{23}$$

whence, from (18), and in complex notation

$$\begin{aligned} V_1 = & -\frac{2m}{e} \frac{\alpha^3}{9p^2} (M + iN)(\cos \xi_1 + i \sin \xi_1)[(\xi \sin \xi + \cos \xi) \\ & + i(\xi \cos \xi - \sin \xi)] \\ & - \frac{2m\alpha^3\beta}{e9p^4} \left[\left\{ \left(\cos \xi_1 - \frac{2}{\xi_1} \sin \xi_1 \right) - i \left(\frac{2}{\xi_1} - \frac{2}{\xi_1} \cos \xi_1 - \sin \xi_1 \right) \right\} \right. \\ & [(\xi \sin \xi + \cos \xi) + i(\xi \cos \xi - \sin \xi)] \\ & \left. - \cos \xi - i(\xi + \tfrac{1}{6}\xi^3 - \sin \xi) \right] + \text{constant}. \end{aligned} \tag{24}$$

With the attainment of (24), the fundamental relation between the alternating-current component J_1 and the alternating-current potential V_1 in the idealized parallel plate diode has been secured. In a more general sense the equation is applicable between any two fictitious parallel planes where one is located at an origin where the boundary conditions for U_0 are satisfied; namely, that the direct-current components of the velocity and acceleration are zero, and the value of the alternating-current velocity at a point, x_1, corresponding to the transit angle, ξ_1, is given by $M \sin pt + N \cos pt$, or by $M + iN$ in complex notation.

Equation (24) contains an additive constant which always appears in potential calculations. This constant disappears when the potential difference is computed. For instance, suppose the potential difference between planes where ξ has the values ξ and ξ', respectively, is desired.

[3] I. Langmuir and Karl T. Compton, "Electrical Discharges in Gases"—Part II, *Rev. Mod. Phys.*, Vol. 3, p. 191; April (1931).

We have

$$V_1 = f(\xi) + \text{constant},$$

$$V_1' = f(\xi') + \text{constant},$$

so that

$$V_1 - V_1' = f(\xi) - f(\xi'). \tag{24-a}$$

Since the potential difference is always required rather than the absolute potential, (24-a) gives the means for applying (24) to actual problems.

III. Application to Diodes

In the application of the fundamental relations to diodes where the thermionic emitter forms the plane located at the origin and the anode coincides with the other plane, the boundary condition is that U_1 shall be zero at the cathode. This means that both M and N are zero and that ξ_1 is also zero. The resulting forms taken by (18) and (24-a), respectively, are as follows:

$$U_1 = -\frac{\beta}{p^2}\left[\left(1 + \cos\xi - \frac{2}{\xi}\sin\xi\right) + i\left(\frac{2}{\xi} - \sin\xi - \frac{2}{\xi}\cos\xi\right)\right], \tag{25}$$

$$V_1 - V_1' = \frac{2m\alpha^3\beta}{e9p^4}\left[(2\cos\xi + \xi\sin\xi - 2) + i(\xi + \tfrac{1}{6}\xi^3 - 2\sin\xi + \xi\cos\xi)\right]. \tag{26}$$

These two equations are identical with those obtained by Benham,[1] and graphs are given in Figs. 1 and 2 showing their variation as a function of the transit angle ξ. In particular, the equivalent impedance between unit areas of the two parallel planes may be found from (26). It must be remembered that the current, A, was assumed positive when directed away from the origin. Hence, we may write

$$Z = -\frac{V_1 - V_1'}{A}. \tag{27}$$

Moreover, the coefficient outside the square brackets in the equation may be expressed more simply when it is realized that the low-frequency internal resistance of a diode is given by the expression

$$r_0 = -\frac{\partial V_0}{\partial J_0}, \tag{28}$$

the minus sign again appearing because of the assumed current direction. Consequently, under the condition of complete space charge,

we have from (22)

$$\frac{2m\alpha^3\beta}{e9p^4} = \frac{12r_0A}{\xi^4} . \tag{29}$$

In addition to the graphs in Figs. 1 and 2 showing the real and imaginary components of impedance and velocity, the graphs shown in Figs. 3 and 4 give their respective magnitudes and phase angles.

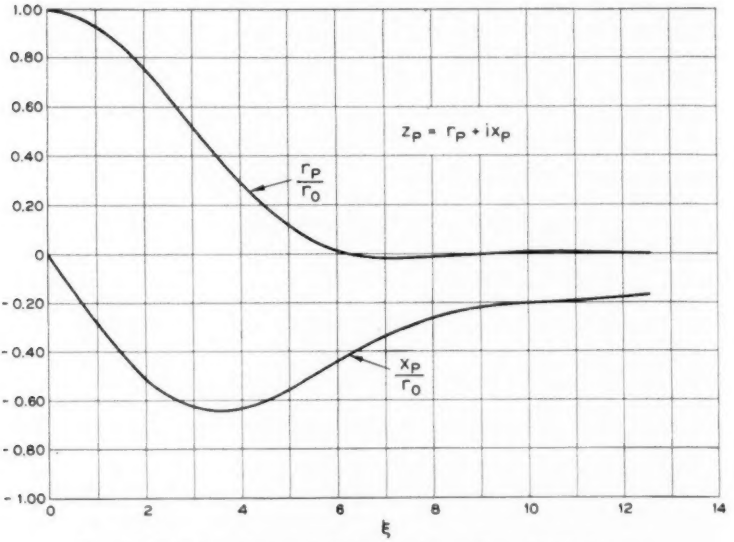


Fig. 1—Plate impedance of diodes or of negative grid triodes as a function of electron transit angle.

The impedance charts show a negative resistance for diodes in the neighborhood of a transit angle, ξ, of 7 radians. The possibility of securing oscillations in this region has been discussed by Benham, so that only a few additional remarks will be made here.

The magnitude of the ratio of reactance to resistance is about 15 when the transit angle is 7 radians. This means that oscillation conditions require an external circuit having a larger ratio of reactance to resistance. On account of the high value of reactance required, a tuned circuit or Lecher-wire system is needed, which would have to operate near an antiresonance point in order to supply the high reactance value. But the resistance component of the external circuit impedance is large at frequencies in the neighborhood of the tuning point, so that the ratio of reactance to resistance is small. Calcula-

tions show that the possibility of securing external circuits having low enough losses to meet the oscillation requirements of most of the diodes which are at present available is not very favorable. The large radio-frequency loss in the filamentary cathodes with which many tubes are supplied is an additional obstacle to be overcome before satisfactory ultra-high-frequency operation of diodes can be expected.

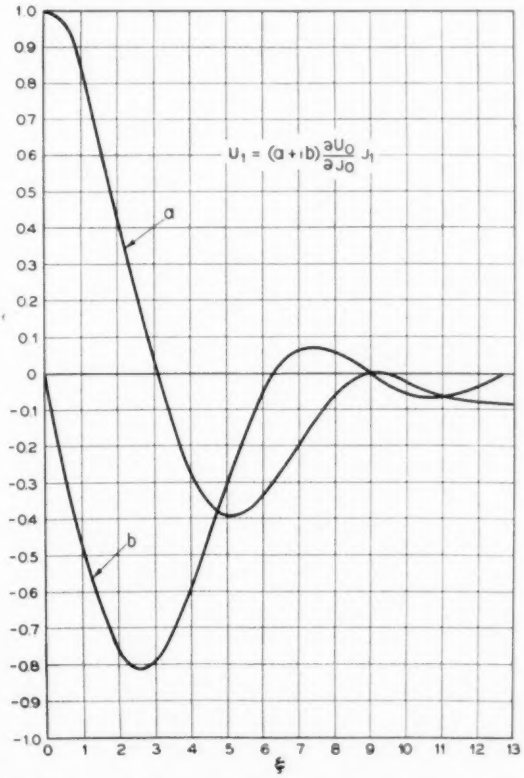


Fig. 2—Electron velocity fluctuation in diodes versus transit angle.

IV. Triodes with Negative Grid and Positive Plate

In the application of the fundamental relations to triodes operating with the grid at a negative potential, the problem becomes more complicated because of the several current paths which exist within the tube. Moreover, the direct-current potential distribution is disturbed in a radical way by the presence of the negative grid. In fact, the

negative grid triode in some respects offers greater theoretical difficulty than does the positive grid triode, which is treated in the next section. However, because of the greater ease in the interpretation of the results in terms which have become familiar through years of use, the negative grid triode is treated first.

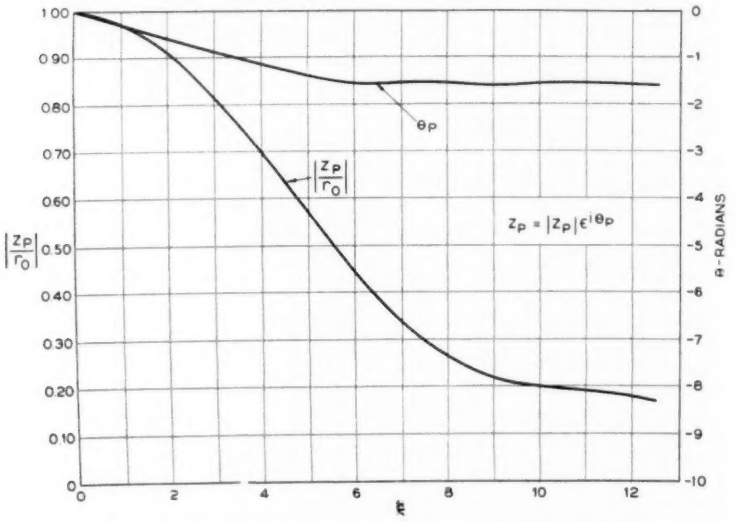


Fig. 3—Magnitude and phase angle of plate impedance of diodes or of negative grid triodes versus transit angle.

In the analysis recourse must be had to approximations and idealizations which allow the theory to fit the practical conditions. In the selection of these, the first thing to notice is that no electrons reach the grid, so that most of the electrostatic force from the grid acts on electrons quite near the cathode, where the charge density is very great. The most prominent effect of a change in grid potential will thus be a change in the velocity of electrons at a point quite near the cathode. It will thus be appropriate to assume as a starting point that the alternating-current velocity at a point x_1, located quite near the cathode is directly proportional to the alternating-current grid potential, V_g, so that we may write,

when $$\xi = \xi_1,$$

$$U_1 = (M + iN) = kV_g. \tag{30}$$

In any event, this relation may be justified if the factor of proportionality, k, be allowed to assume complex values, and ξ_1 is not taken too near the origin. Actually, the electron-free space surrounding the grid wires, and the fact that the electric intensity at a point midway between any two of the wires is directed perpendicularly to the plane of the grid, gives us more confidence in extending the approximation, so that k will be regarded as real, and ξ_1 will be taken very small.

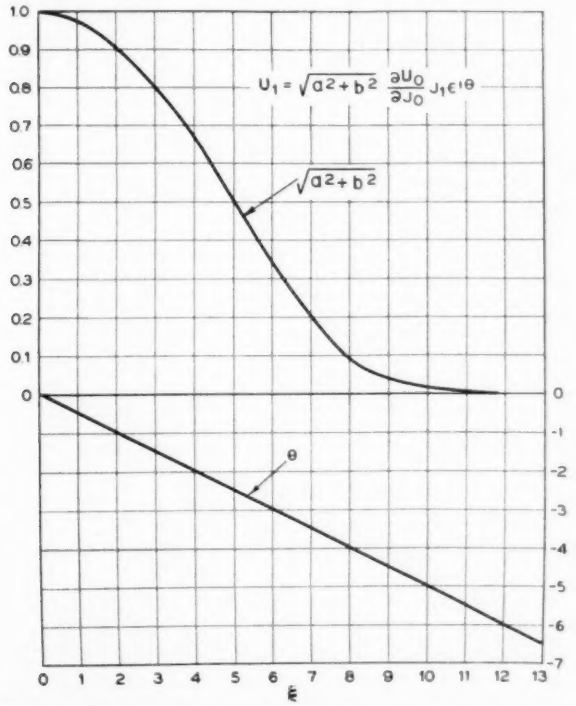


Fig. 4—Magnitude and phase angle of electron velocity fluctuation in diodes versus transit angle.

Equation (24) may, therefore, be applied under the conditions that $\xi_1 \to 0$, and gives the following for the potential difference between plate and cathode:

$$V_p = -\frac{12 r_0 A}{\xi^4}\left[(\xi \sin \xi + 2\cos\xi - 2) + i(\xi + \tfrac{1}{6}\xi^3 - 2\sin\xi + \xi\cos\xi)\right.$$
$$\left. -(M+iN)\frac{b^2}{\beta}\left[(\xi\sin\xi + \cos\xi - 1) - i(\sin\xi - \xi\cos\xi)\right]\right]. \quad (31)$$

This equation may be written in condensed form with the aid of (30)

$$V_p = J_1(r + ix) - V_g(\mu + i\nu), \tag{32}$$

where

$$\left.\begin{aligned} r &= -\frac{12r_0}{\xi^4}(\xi \sin \xi + 2\cos \xi - 2), \\ x &= -\frac{12r_0}{\xi^4}(\xi + \tfrac{1}{6}\xi^3 - 2\sin \xi + \xi \cos \xi), \\ \mu &= \frac{2\mu_0}{\xi^2}(\xi \sin \xi + \cos \xi - 1), \\ \nu &= \frac{2\mu_0}{\xi^2}(\xi \cos \xi - \sin \xi). \end{aligned}\right\} \tag{33}$$

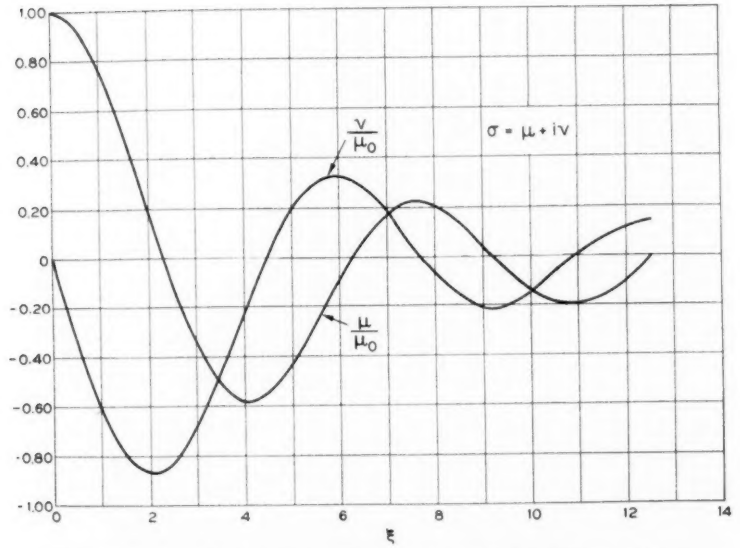


Fig. 5—Real and imaginary components of complex amplification factor of negative grid triodes versus transit angle.

The significance of (32) is at once apparent when it is compared with the classical form of the equation representing the alternating-current plate voltage, namely,

$$V_p = I_p r_0 - \mu V_g.$$

The plate resistance r_0 has now become complex as likewise has the amplification factor μ. Values of the plate impedance

$$z_p = r + ix$$

are the same as those obtained for the diode and are plotted in Figs. 1 and 3. Values of the amplification factor

$$\sigma = \mu + i\nu$$

are shown in Figs. 5 and 6.

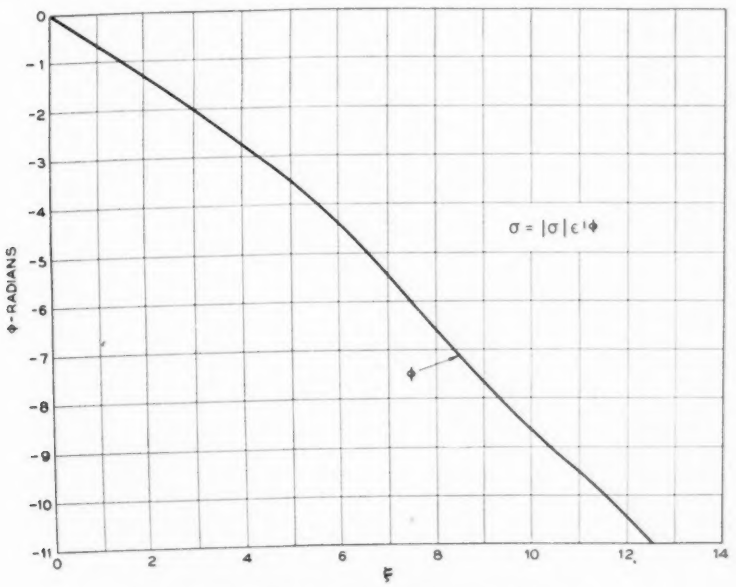


Fig. 6—Phase angle of amplification factor of negative grid triodes versus transit angle.

It is evident that radical changes in the phase angles existing between the grid voltage and plate current are present when the transit time becomes appreciable in comparison with the period of the applied electromotive force. The plate impedance decreases in magnitude as also does the magnitude of the amplification factor. However, the ratio of the two, namely, σ/z, maintains a fairly constant magnitude as shown in Fig. 7, whose phase angle nevertheless rotates continually in a negative direction becoming equal to 3 radians when ξ is 2π.

The interelectrode capacity between cathode and plate is included in the fundamental relations here employed. This inclusion exhibits one important difference between (32) and the classical case. At low frequencies, the equivalent circuit represented by (32) degenerates into that shown on Fig. 8. The capacity branch exists in parallel with the

resistive branch and they are both in series with the effective generator σe_g, whereas in the classical picture the capacity branch shunts the effective generator and plate resistance which are in series with each other. Practically the difference between the two equivalent circuits is negligible except at extremely high frequencies. The following physical viewpoint supports the newer picture.

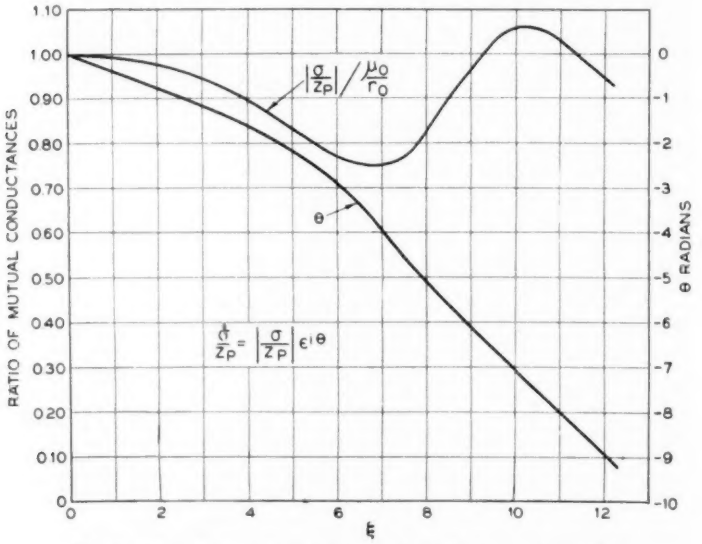


Fig. 7—Magnitude of complex mutual conductance of negative grid triodes versus transit angle.

As pointed out, the action of the grid is exerted mostly on the region of dense space charge existing very near the cathode and variations in the grid potential act on the velocities of the emerging electrons, thus producing the equivalent generator of the plate circuit. The plate current consists of conduction and displacement components whose sum is the same at all points in the cathode-plate path. Near the cathode, the conduction component comprises the whole current because of the high charge density and the effective generator acts in series with this current and hence in series with the path of the displacement current into which the character of the total current gradually changes as the plate is approached.

Strictly speaking, the equivalent circuit corresponding to (32) exists, not between the plate and cathode, but between the plate and the potential minimum near the cathode which is caused by the finite

velocities with which electrons are emitted from the cathode. Practically, the difference is negligible except at extremely high frequencies. Since the impedance between the cathode and potential minimum is small compared to the plate impedance, its effect is merely to add a loss to the system which increases with frequency since the plate impedance approaches a capacity as the frequency approaches infinity.

The grid cathode path presents less difficulty, although a somewhat less rigorous treatment is given here. As pointed out, the force from the grid acts on the high charge density region existing near the potential minimum. The impedance between cathode and grid, therefore, consists of two parts in series; namely, capacity between grid and potential minimum and impedance between potential minimum and cathode, the latter part of this impedance being common both to plate- and grid-current paths.

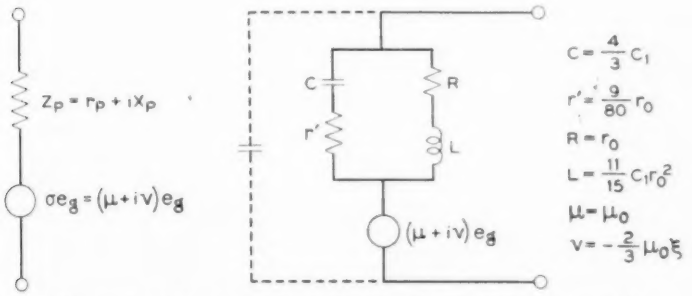


Fig. 8—Equivalent network of plate-cathode path of negative grid triodes for transit angles less than 0.3 radian.

If we were to connect the grid and cathode terminals of such a triode to a capacity bridge and measure the capacity existing there when the tube was cold and when the cathode was heated, we should find that the capacity would exhibit a slight increase in the latter case. The reason for this increase may best be explained by noting that in the cold condition the electrostatic force from the grid is exerted on the cathode itself, whereas in the heated state, the force acts on the electrons near the potential minimum, thus resulting in an increased capacity in series with a resistive component.

In some measurements of the losses in coils which were made at a frequency of 18 megacycles, J. G. Chaffee of the Bell Telephone Laboratories has found that a loss existed between grid and cathode of vacuum tubes which was much greater than can be accounted for by any of the dielectrics used and which was present only when the tube

filament was hot. This loss increased with frequency in the manner characteristic of that of the capacity-resistance combination between cathode and grid which was described above. Present indications are that, at least in part, the loss may be ascribed to the resistance existing between the cathode and the region of potential minimum.

Of the three current paths through the tube, one more still remains to be considered. This is the grid-plate path. The relations involved here are more readily seen by considering first a low-frequency example. Here the electron stream passes through the spaces between grid wires, afterward diverging as the plate is approached. Electrostatic force from the grid acts not only on the plate but also on the electrons in the space between. It is evident, then, that the path which, when the cathode was cold, constituted a pure capacity changes into an effective capacity different from the original in combination with a resistive component. The losses would be expected to increase with frequency just as they did in the grid-cathode type. The change in grid-plate impedance is particularly noticeable when it is attempted to adjust balanced or neutralized amplifier circuits with the filament cold, in which case the balance is disturbed when the cathode is heated.

As yet, no accurate expression for this grid-plate impedance has been obtained, either at the low frequencies where transit times are negligible or at the higher frequencies now particularly under investigation. The reason for this lies in the repelling force on the electron stream of the negative grid so that the assumption of current flow in straight parallel lines is not valid in so far as current from the grid to the plate is involved.

It has been shown that both the cathode-grid path and the grid-plate path contain resistive components with corresponding losses which increase with increase of frequency. This loss may be cited as a reason why triodes with negative grids cease to oscillate at the higher frequencies. If it were not for these losses, external circuits could be attached to the tube having such phase relations as to satisfy oscillation conditions, so that the negative grid triode could be utilized in the range which is now covered by the triode with positive grid.

V. Triodes with Positive Grid and Slightly Positive Plate

When the grid of a three-element tube is operated at a high positive potential with respect both to cathode and plate, electrons are attracted toward the grid, and the majority of them are captured on their first transit. Those which pass through the mesh and journey toward the plate will be captured by the plate if its potential is sufficiently positive with respect to the cathode.

In general, space-charge conditions existing between grid and plate are quite complicated. An analysis has been made by Tonks [4] which indicates several distinct classes of space-charge distribution which are possible. In the first place so few electrons may pass the grid mesh that no appreciable space charge is set up between there and the plate. In this instance a positive plate will trap them all, whereas a negative plate will return them all toward the grid. Second, with a fixed positive plate potential an increase in the number of electrons which pass the grid mesh will result in a depression of the potential distribution as illustrated at (*a*) by the curves in Fig. 9. This depression will con-

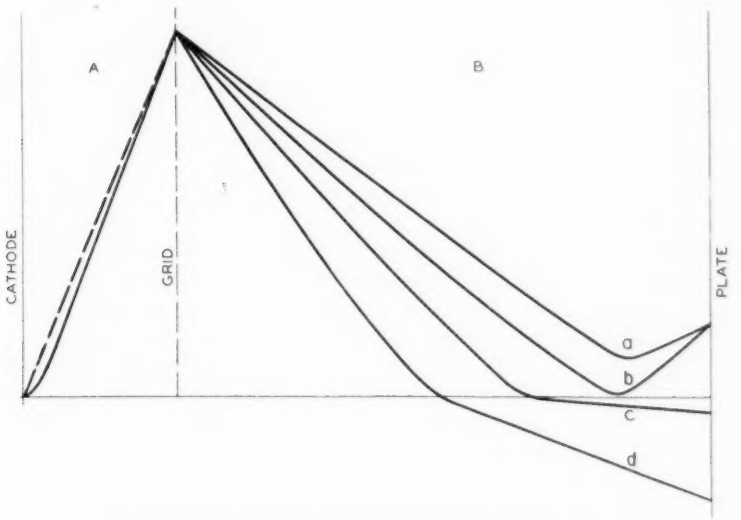


Fig. 9—Potential distributions in positive grid triodes.

tinue to increase until a potential minimum is formed. When this potential minimum becomes nearly the same as that of the cathode, either of several things may occur. If the minimum is just above the cathode potential, all electrons will pass that point and eventually reach the plate. However, an extremely small increase in the number of electrons will cause the potential minimum to become equal to the cathode potential. When this happens some of the electrons will be turned back and travel again toward the grid. These will increase the charge density existing and, therefore, cause a further depression in the potential resulting in a mathematical discontinuity so that the

[4] L. Tonks, "Space Charge as a Cause of Negative Resistance in a Triode and Its Bearing on Short-Wave Generation," *Phys. Rev.*, Vol. 30, p. 501; October (1927).

curve of the potential suddenly changes its shape with a resulting change in plate current. Again, the plate may be operated at a negative potential. In this case, none of the electrons will reach it and the potential distribution curves have the character illustrated at (*c*) and (*d*) in Fig. 9.

In attempting to apply the fundamental relations to this grid-plate region, we must choose our origin at a point where the potential distribution curve touches the zero axis and is tangent to it. Whenever such a point exists, the relations may be applied as described below. Even when this condition does not exist inside the vacuum tube, there may be a virtual cathode existing outside of the plate.

Whenever all of the electrons passing the grid reach the plate the general equations may be applied in a straightforward manner with the origin taken at the virtual cathode. Whenever some of the electrons are turned back toward the grid, slightly different equations are required, although they may be applied in the same manner. These modified equations will be derived and discussed after the application of the equations already derived has been made to the case where all of the electrons reach the plate.

Choosing the origin for this latter case at the point of zero potential or virtual cathode, we can compute the impedance between the grid-plane and the virtual cathode when we know the alternating-current velocities with which the electrons pass through the grid-plane. This has been found for the condition of complete space charge between cathode and grid and was given by (25). Likewise, it can be found on the supposition that no space charge exists in the cathode-grid region and the result will be calculated later. Thus, two limiting cases are available for numerical application.

In order to prevent confusion for the grid—virtual-cathode region where the electron flow is toward the origin rather than away from it, as was assumed in the derivation of the fundamental relations, it will be convenient to change the symbol for transit angle from ξ to $-\zeta$. This will automatically take care of all algebraic signs, currents and velocities now being considered positive when directed towards the origin.

Since we are computing the impedance between an origin at the virtual cathode and the grid plane we may apply (24) to find the potential difference, getting

$$V_1 - V_1' = -\left(\frac{2m}{e}\frac{\alpha^3}{9p^2}\right)(M + iN)[(1 - \cos\zeta) - i(\zeta - \sin\zeta)]$$
$$- i\left(\frac{2m}{e}\frac{\alpha^3\beta}{9p^4}\right)\left(\frac{1}{6}\zeta^3 - \frac{4}{\zeta}(1 - \cos\zeta) + 2\sin\zeta\right), \quad (34)$$

where V_1' is the potential at the virtual cathode.

This relation is of the form

$$V_g - V_p = -(M+iN)\left(\frac{2m}{e}\frac{\alpha^3}{9p^2}\right)[(1-\cos\zeta)-i(\zeta-\sin\zeta)]+J_pZ_p, \quad (35)$$

where J_p is the plate current, and Z_p is the effective impedance:

$$Z_p = -i\left(\frac{2m}{e}\frac{\alpha^3\beta}{9p^4A}\right)\left(\frac{1}{6}\zeta^3-\frac{4}{\zeta}(1-\cos\zeta)+2\sin\zeta\right). \quad (36)$$

In terms of the cold capacity C_1 between plate and grid plane this becomes

$$Z_p = -\frac{i}{pC_1}\frac{6}{\zeta^3}\left[\frac{1}{6}\zeta^3-\frac{4}{\zeta}(1-\cos\zeta)+2\sin\zeta\right],$$

which is plotted in Fig. 10.

The form of (35) shows that the equivalent network between the plane of the grid and the plate may be represented by an equivalent generator acting in series with the impedance, Z_p. This is evidenced by the fact that the velocity $M + iN$ with which the electrons pass the grid, may be expressed in terms of the grid potential V_g by means of conditions between the grid and cathode. When complete space charge exists near the cathode, these conditions are expressed by (25) and (26). On the other hand, tubes with positive grid are sometimes operated with inappreciable space charge between grid and cathode. In this event, a similar analysis leads to values for the alternating-current velocity and potential at the grid as follows:

$$U_1 = M + iN = -\frac{\beta}{p^2}\left[\left(\frac{\eta-\sin\eta}{\eta}\right)+i\left(\frac{1-\cos\eta}{\eta}\right)\right], \quad (37)$$

$$V_g = i\frac{4\pi x}{p}A = \frac{iA}{pC}, \quad (38)$$

where η is the transit angle in the absence of space charge, and C is the electrostatic capacity between unit area of cathode and of grid plane. The right-hand side of (38) does not contain a minus sign because of the assumed current direction which is away from the cathode, as is also the convention employed in (25) and (26) where the electron charge e is a positive number.

The relations given by (35) allow the potential difference between grid and plate to be determined in terms of the total current flowing to the plate, and the total current flowing from the cathode, which ap-

pears in the velocity factor $M + iN$. In the usual case some of the alternating current flows to the grid wires and is returned through an external circuit connected to the grid. If the impedance between grid and plate is desired it is necessary to find the relation which this grid current bears to the total cathode and plate currents and to the alternating-current potentials. The calculations involved are extremely complicated because the assumption of current flow in straight lines between parallel planes is far from representing the actual conditions

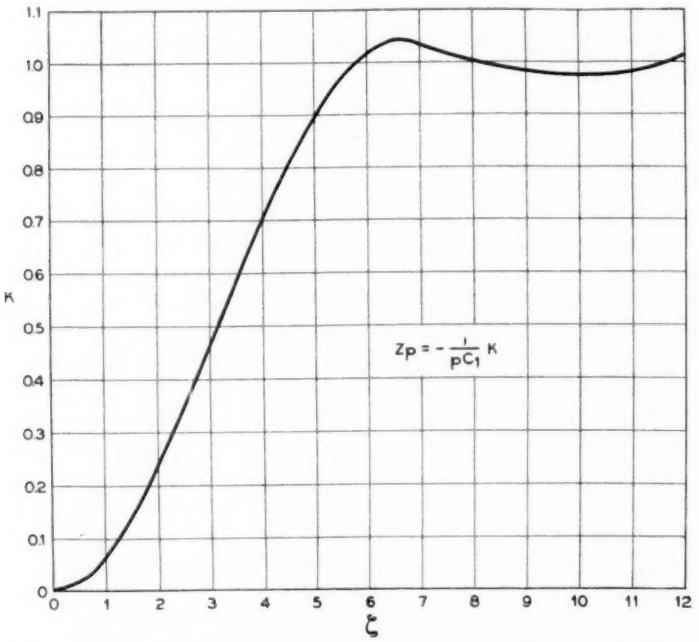


Fig. 10—Plate impedance of positive grid triodes with slightly positive plate.

in the immediate neighborhood of the grid wires. Rather than attempting an analysis of these conditions at the present time, we shall content ourselves with results already obtained, since they are applicable to the special case, which can be realized approximately in experiment, where the grid is connected to a radio-frequency choke coil of sufficiently good characteristics to prevent it from carrying away any alternating current. For this special case the current J_1 is the same both in the cathode region and in the plate region, and all encumbering assumptions involving different paths for the conduction and displace-

ment components of the current in the neighborhood of the grid wires have been done away with.

The application of the equations to this special case is dealt with in the section of this paper devoted to positive grid oscillators. Before these oscillators can be treated comprehensively, a further extension of fundamental theory is necessary. This extension comes about because positive grid oscillators are often operated with a slightly negative potential applied to the plate.

VI. Triodes with Positive Grid and Negative Plate[5]

When consideration is directed to tubes operating with positive grid but negative plate, the fundamental underlying theory must again be investigated. The reason for this lies in the fact that all electrons which penetrate through the meshes of the grid are turned back before they reach the plate, so that in the grid-plate space there are two streams of electrons moving in opposite directions. The effect of this double value for the velocity may readily be calculated in so far as direct-current components, only, are concerned. We have merely to note that the charge density is double the value which it would have in the presence of those electrons which are moving in one direction, only, so that the correct relations are obtained from the equations already derived by taking twice the value of direct current in one direction.

When alternating-current components are considered, however, matters are more complicated, but not difficult. To see what the actual relations are, let there be two possible values at any point for the instantaneous velocity, and call these two values U_a and U_b, respectively. Then the relation between force and acceleration becomes

$$\frac{eE}{m} = \frac{dU_a}{dt} = \frac{dU_b}{dt}. \tag{39}$$

Hence, at a given value of x we have by integration

$$U_a = U_b + \text{constant}.$$

But, when both values of velocity are separated into their components according to (5) we have from (39)

$$U_{a0} + U_{a1} + \cdots = U_{b0} + U_{b1} + \cdots + \text{constant}.$$

[5] Since the publication of this paper in the *Proceedings* of the *Institute of Radio Engineers*, several questions have been raised regarding the treatment presented in sections VI and VII. These are being investigated and will form the basis of another paper.

By equating corresponding terms, we find

$$\left.\begin{aligned} U_{a0} &= U_{b0} + \text{constant}, \\ U_{a1} &= U_{b1}, \\ U_{a2} &= U_{b2}, \text{ etc.} \end{aligned}\right\} \tag{40}$$

The first of these equations is trivial when the boundary conditions are inserted, for then it appears that $U_{a0} = -U_{b0}$ and the equation merely states that at a given value of x the direct-current velocity component is not a function of time.

The second equation is much more enlightening and tells us that although two values of the direct-current velocity may be present, nevertheless there is only a single value for the alternating-current component. The same conclusion holds for the higher order velocity components. This conclusion supplies the key for the solution of the general equations when applied to the stream of electrons moving in both directions between the grid and plate of the tube.

In general, the total current may be written

$$J = P_a U_a + P_b U_b + \frac{1}{4\pi}\frac{\partial E}{\partial t}. \tag{41}$$

If Σ is the total area of each of the electrode planes and

$$\Sigma = a + b,$$

where a and b are constants to be defined later, (41) may be written as follows:

$$J\Sigma = \left(P_a U_a \frac{\Sigma}{a} + \frac{1}{4\pi}\frac{\partial E}{\partial t}\right) a + \left(P_b U_b \frac{\Sigma}{b} + \frac{1}{4\pi}\frac{\partial E}{\partial t}\right) b. \tag{42}$$

In this expression, the two streams of current are clearly separated if a and b are taken so that [6]

$$P_a \frac{\Sigma}{a} = P \qquad \text{and} \qquad P_b \frac{\Sigma}{b} = P, \tag{43}$$

where P is the total charge density, equal to the sum of P_a and P_b.

The total current may now be expressed in terms of velocities, only, giving similarly to the transition from (1) to (3),

$$4\pi \frac{e}{m} J\Sigma = a\left(U_a \frac{\partial}{\partial x} + \frac{\partial}{\partial t}\right)^2 U_a + b\left(U_b \frac{\partial}{\partial x} + \frac{\partial}{\partial t}\right)^2 U_b. \tag{44}$$

[6] A more rigorous analysis, involving mean values of the motions of individual electrons, leads to the same result.

When U_a and U_b are each separated into their components according to (5), so that (44) may be resolved into a system of equations, we have for the first two equations, analogous to (6) and (7),

$$4\pi \frac{e}{m} J_0 \Sigma = (a - b) \left[U_0 \frac{\partial}{\partial x} \left(U_0 \frac{\partial U_0}{\partial x} \right) \right] \tag{45}$$

and

$$\beta\Sigma = (a+b) \left[U_0 \frac{\partial}{\partial x} \left(U_0 \frac{\partial U_1}{\partial x} + U_1 \frac{\partial U_0}{\partial x} \right) + \frac{\partial^2 U_1}{\partial t^2} + U_1 \frac{\partial}{\partial x} \left(U_0 \frac{\partial U_0}{\partial x} \right) \right] + (a-b) \left[U_0 \frac{\partial}{\partial x} \left(\frac{\partial U_1}{\partial t} \right) + \frac{\partial}{\partial t} \left(U_0 \frac{\partial U_1}{\partial x} + U_1 \frac{\partial U_0}{\partial x} \right) \right], \tag{46}$$

where the components of U_b have been expressed in terms of those of U_a by means of (40) and the relation that $U_{b0} = - U_{a0}$.

The solution of (45) is, as before,

$$U_0 = \alpha x^{2/3}, \tag{47}$$

where

$$\alpha = \left(18\pi \frac{e}{m} J_{0a} \frac{\Sigma}{a} \right)^{1/3}.$$

Before attempting to solve (46) we make a change of variable as in (10), writing

$$\xi = \frac{3p}{\alpha} x^{1/3} \qquad \text{and} \qquad U_1 = \frac{\omega}{\xi}.$$

This gives from (46)

$$\beta\Sigma = \left[\frac{p^2}{\xi} \frac{\partial^2 \omega}{\partial \xi^2} + \frac{1}{\xi} \frac{\partial^2 \omega}{\partial t^2} \right] \Sigma + (a - b) \left(\frac{2p}{\xi} \frac{\partial^2 \omega}{\partial \xi \partial t} \right). \tag{48}$$

In finding a solution for this, we shall restrict ourselves to the case where all of the electrons turn back at the virtual cathode, so that $a = b$ and therefore the last term of (48) vanishes. The solution of the remaining equation is then,

$$U_1 = - \frac{\beta}{p^2} \left[\sin pt + \frac{1}{\xi} F_1(i\xi + pt) + \frac{1}{\xi} F_2(i\xi - pt) \right], \tag{49}$$

which is analogous to (12).

Again, assuming the two arbitrary functions to have the form,

$$\left.\begin{aligned} F_1(i\xi + pt) &= a \sin (i\xi + pt) + b \cos (i\xi + pt), \\ F_2(i\xi - pt) &= c \sin (i\xi - pt) + d \cos (i\xi - pt), \end{aligned}\right\} \tag{50}$$

and inserting the boundary conditions, (14) and (15), we have, in complex form,

$$U_1 = (M + iN)\frac{\xi_1}{\xi}\frac{\sinh \xi}{\sinh \xi_1} - \frac{\beta}{p^2}\left(1 - \frac{\xi_1}{\xi}\frac{\sinh \xi}{\sinh \xi_1}\right), \tag{51}$$

which is a simpler equation than its analogue (18). The potential is obtained as in (24) giving,

$$\begin{aligned} V_1 = &-\left(\frac{\alpha^3 m}{9p^2 e}\right)(M + iN)\frac{\xi_1}{\sinh \xi_1}[\xi \sinh \xi + i(\xi \cosh \xi - \sinh \xi)] \\ &+ \frac{\alpha^3 m\beta}{9p^4 e}\left[\left(\xi^2 - \frac{\xi_1 \xi \sinh \xi}{\sinh \xi_1}\right)\right. \\ &\left. + i\left(\frac{\xi^3}{3} - \frac{\xi_1}{\sinh \xi_1}(\xi \cosh \xi - \sinh \xi)\right)\right] + \text{constant}. \end{aligned} \tag{52}$$

The alternating-current potential difference between the grid and the virtual cathode where all of the electrons are turned back may be obtained immediately from (52). As before, the variable ζ will be substituted for ξ to show that the grid-plate region is considered, and currents and velocities will be considered positive when directed towards the origin at the virtual cathode. Thus, from (52)

$$\begin{aligned} V_g - V_p = &-\frac{\alpha^3 m}{9p^2 e}(M + iN)[\zeta^2 + i(\zeta^2 \coth \zeta - \zeta)] \\ &+ \frac{\alpha^3 m\beta}{9p^4 e} i\left[\frac{\zeta^3}{3} - \zeta^2 \coth \zeta + \zeta\right]. \end{aligned} \tag{53}$$

The velocity, $(M + iN)$ may be expressed in terms of the alternating-current grid potential, V_g, so that the path between grid plane and virtual cathode may be represented by an effective generator in series with an impedance, as was done in (34), (35), and (36).

VII. Oscillation Properties of Positive Grid Triodes [5]

The oscillation properties of the positive grid triode are next to be investigated. In the usual experimental procedure, an external high-frequency circuit is connected between the grid and the plate of the tube. It is unfortunate that this particular arrangement greatly complicates the theoretical relations. Accordingly, a slightly modified experimental set-up will be considered. This modification consists in connecting the external circuit between the cathode and plate of the tube, rather than between grid and plate. Experimental tests have shown that the modified circuit exhibits the same general phenomena

[5] Loc. cit.

as the more usual one, the difference being mainly one of mechanical convenience in securing low-loss leads between the tube and the external circuit.

The modified circuit, then, will be employed for analysis, and the assumption will be made that the necessary direct-current connections are made through chokes which are sufficiently good so that it may be considered that no external high-frequency impedance is connected between either the grid and the plate, or between the cathode and the grid.

It is easy to see that under these conditions there can be no high-frequency current carried away by the grid. It follows that for plane-parallel structures, the alternating-current density, J_1, will be the same both in the cathode-grid region and in the grid-plate region. The arrangement thus reduces the problem to the consideration of the single current, J_1, and the resulting potential difference between cathode and plate.

There are several possible combinations of direct-current biasing potentials. For the first of these, the plate will be supposed to be biased at a potential sufficiently positive to collect all electrons which are not captured by the grid on their first transit. Complete space charge will be assumed both in the cathode region and in the plate region.

Under these conditions, we have the grid-cathode potential difference given by (26) and the grid-plate potential difference given by (35), where the velocity, $M + iN$, is given by (25). We can write,

$$\begin{aligned} V_p - V_c &= (V_p - V_g) + (V_g - V_c) \\ &= -\,[\text{Eq. } 35] + [\text{Eq. } 26]. \end{aligned} \tag{55}$$

It will be remembered that the current was assumed to be positive in (26) when directed away from the origin, and positive in (35) when directed toward the origin. Therefore, since the same current exists in both regions, and they are joined together at the grid, the sign of the current J_1 remains the same in both (35) and (26), its direction being from cathode to plate. The impedance looking into the cathode-plate terminals may be obtained from (55) by dividing by the amplitude A of J_1 and reversing the sign of the result to correspond to a current from plate to cathode. Letting

$$Z_0 = R_0 + iX_0 \tag{56}$$

represent the impedance looking into the cathode-plate terminals, we can write the result as follows

$$R_0 = -\frac{12r_0}{\zeta^4}\left[\left(1+\cos\eta-\frac{2}{\eta}\sin\eta\right)(1-\cos\zeta)\right.$$
$$-\left(\frac{2}{\eta}-\sin\eta-\frac{2}{\eta}\cos\eta\right)(\zeta-\sin\zeta)$$
$$\left.+(2\cos\eta\sin\eta-2)\right], \tag{57}$$

$$X_0 = -\frac{12r_0}{\zeta^4}\left\{\left(1+\cos\eta-\frac{2}{\eta}\sin\eta\right)(\zeta-\sin\zeta)\right.$$
$$+\left(\frac{2}{\eta}-\sin\eta-\frac{2}{\eta}\cos\eta\right)(1-\cos\zeta)$$
$$+\left[\frac{1}{6}\zeta^3-\frac{4}{\zeta}(1-\cos\zeta)+2\sin\zeta\right]$$
$$\left.+\left[\eta+\frac{1}{6}\eta^3-2\sin\eta+\eta\cos\eta\right]\right\}, \tag{58}$$

where η is the transit angle from cathode to grid, ζ is the transit angle from grid to virtual cathode at the plate, and r_0 is the zero-frequency resistance which would be present in a diode having the grid-plate dimensions, and the same operating direct-current voltages and current densities which occur in the grid-plate region of the triode under consideration.

Fig. 11 shows graphically the relation between R_0 and X_0 for a wide frequency range, in terms of the reference resistance, r_0. Curve A is drawn for the hypothetical condition that $\eta = \zeta$, so that the tube is exactly symmetrical about the grid. Actually such a condition could not be attained, since the grid captures some of the electrons, leaving fewer for producing space charge near the plate. The grid-plate dimension would accordingly have to be increased in order to secure the space charge, but this would cause the transit angle ζ to become larger than η. However, despite the fact that it does not correspond to a physically realizable condition, curve A is nevertheless of use in indicating the limit which is approached as the grid capture fraction is made smaller and smaller.

Curves B and C correspond to values of grid-plate transit angle equal respectively to two and three times the cathode-grid transit angle. Both these curves represent conditions which may readily be obtained experimentally, and indeed, curves lying much closer to A than does the curve B may be secured. For example, the general relation for the ratio of the transit angles in terms of the direct currents J_a and J_b in the cathode and in the plate region, respectively, when

complete space charge exists in both regions, is,

$$\frac{\zeta}{\eta} = \sqrt{\frac{J_a}{J_b}}.$$

Suppose that the grid captured half of the electrons. Then the ratio of transit angles would be 1.41. This would result in a curve lying between A and B in Fig. 11.

The numbers, $\pi/2$, π, and so forth, which are attached to the curves in Fig. 11 show the values of the grid-plate transit angle, ζ, which correspond to the points indicated.

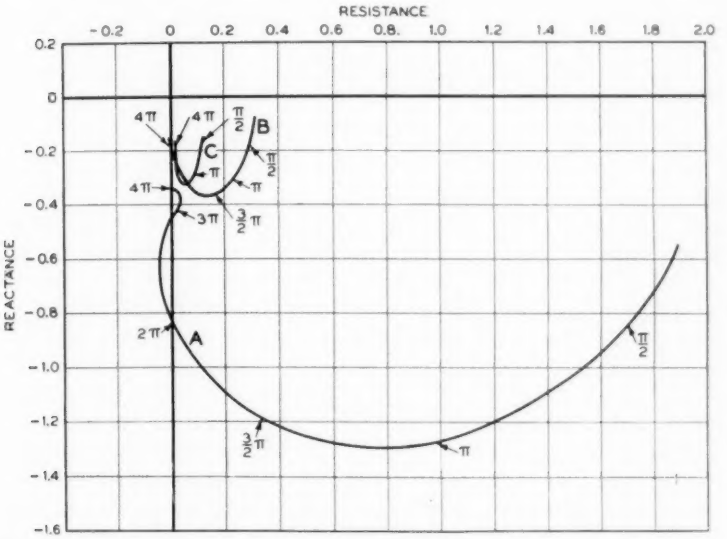


Fig. 11—$R_0 - X_0$ diagram for positive grid, slightly positive plate triode with cathode space charge.

Curve A, $\eta = \zeta$
Curve B, $\eta = 1/2\zeta$
Curve C, $\eta = 1/3\zeta$

We now come to the problem of obtaining information about the oscillation properties of a tube from a set of curves such as those shown in Fig. 11. In a very loose way, and without proof we may state the results of an extension of Nyquist's [7] rule as follows:

If an $R - X$ diagram, which in general may include negative as well as positive frequencies, encircles the origin in a clockwise direc-

[7] H. Nyquist, "Regeneration Theory," *Bell Sys. Tech. Jour.*, Vol. 11, p. 126; January (1932).

tion, then the system represented by the diagram will oscillate when the terminals between which the impedance was measured are connected together.

Verification of this rule, together with further extension to more general cases are expected to be discussed in a subsequent paper. For the present, its validity will have to be accepted on faith, but with the assurance that the applications employed in this discussion are readily capable of demonstration.

Returning to consideration of the positive grid triode with complete space charge on both sides of the grid, and a slightly positive plate, whose $R - X$ diagram is given in Fig. 11, we see at once that the diagram does not encircle the origin as it stands. Of course only positive values of frequency are included in the curves as they are shown. The inclusion of negative frequencies (never mind their physical meaning) would produce a curve which would be the image of the curve shown, a reflecting mirror being regarded as a plane perpendicular to the paper, and containing the R-axis. The curve A, for instance, would have its part corresponding to negative frequencies lying above the R-axis and forming an image of the part lying below. This is shown by the dotted curve in Fig. 12.

It is obvious that the curve of Fig. 12 will encircle the origin or not depending on what happens at infinite frequencies. However, the slightest amount of resistance in the leads to the tube will be sufficient to move the curve to the right and thus exclude the origin. This means that no oscillations would be obtained if an alternating-current short were placed between plate and cathode. The result, although in accord with experiment, is not particularly useful. The important thing is to find whether the curve can be modified by the addition of a simple electrical circuit in such a way that the origin of the resulting $R - X$ diagram for the combination of tube and circuit is encircled in a clockwise direction.

Suppose that a simple inductance is connected in series with the plate lead, and the impedance diagram of the series combination of tube and inductance is plotted. For this arrangement, the $R - X$ diagram of Fig. 12 would be modified as shown in Fig. 13. Here the part of the curve corresponding to negative values of resistance has been pushed upward until the origin is enclosed within a loop which encircles it in a clockwise direction. It is therefore to be expected that oscillations will result. As to their frequency, we can say that the grid-plate transit angle must be at least as great as 2π for this particular example. This follows by supposing a certain amount of resistance to be added in series with the circuit. The effect of this resistance will

be to move the curves on Fig. 13 bodily to the right. The lowest frequency which will just allow the origin to be included within the loop when the series resistance is reduced to zero and the inductance is adjusted, corresponds to a grid-plate transit angle of 2π.

It must be remembered that the foregoing details apply only to curve *A* of Fig. 11, and it has already been pointed out that curve *A*

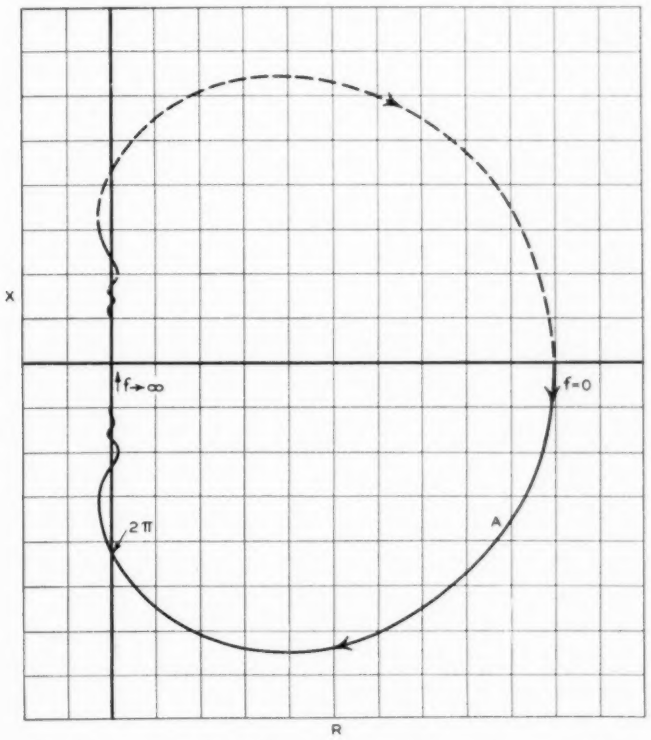


Fig. 12—Curve *A* of Fig. 11 together with the image corresponding to negative frequencies.

represents a limit which can be approached in practice, only as the grid capture fraction is made smaller and smaller. Curve *B* can well be duplicated in experiment. For this case, the lowest frequency at which oscillations may be expected is much higher than before, since the transit angle must be equal to 4π before the resistance becomes negative. Actually, conditions intermediate between the two curves may be realized, so that from a practical standpoint the transit angle

must be in the neighborhood of 3π before we may expect to secure oscillations.

This would correspond to a frequency somewhat higher than is often associated with this type of oscillation. It must be remembered however, that the particular case considered was that of a tube with its plate at a slightly positive potential, whereas the majority of the experimental frequency observations were made with the plate either slightly negative, or, if positive, adjusted so that a virtual cathode was formed inside the tube, and many of the electrons were turned back

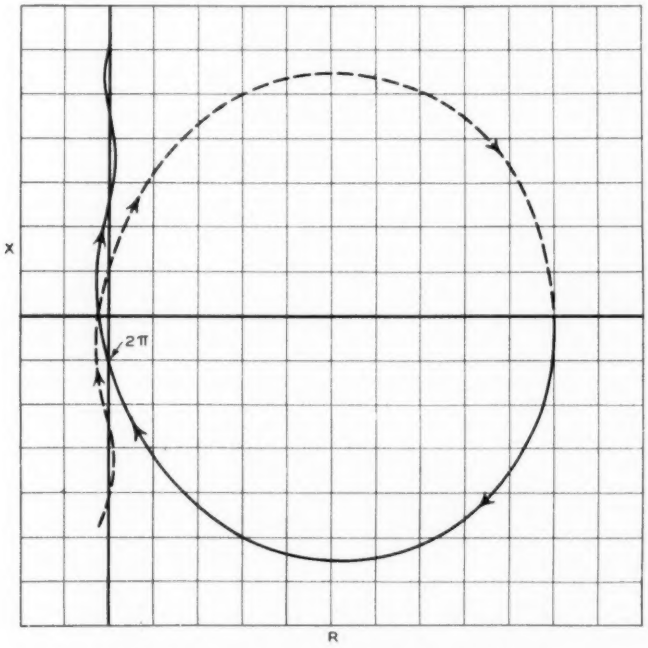


Fig. 13—Modification of Fig. 12 produced by added inductance.

before they reached the plate. The curves of Fig. 11 do not apply to these cases.

Therefore, let us see what happens when the plate is operated at a negative potential so that all of the electrons are turned back before they reach it. At the outset, it should be remarked that this condition does not prohibit the presence of direct-current plate current *after the oscillations have built up to a finite amplitude.* The analysis applies to the requirements for the starting of the oscillations, only, so that

if the plate fluctuates in potential by a very small amount, as it does for incipient oscillations, and hence does not become positive during the alternating-current alternation, then no direct-current plate current can occur when the plate is biased negatively. After oscillations have built up to an appreciable amplitude, the presence of plate current is not only possible, but is in fact to be expected.

We have at hand the mathematical tools with which to compute our $R - X$ diagram for the negative plate triode with complete space charge near the cathode. Thus, instead of substituting (35) in (55) we must substitute (53). Since complete space charge is still postulated near the cathode, (26) and (25) are still applicable. The result is:

$$R_0 = -\frac{12r_0}{\zeta^4}\left[\left(1 + \cos\eta - \frac{2}{\eta}\sin\eta\right)\zeta^2 - \left(\frac{2}{\eta} - \sin\eta - \frac{2}{\eta}\cos\eta\right)(\zeta^2\coth\zeta - \zeta) + (2\cos\eta + \eta\sin\eta - 2)\right], \tag{59}$$

$$X_0 = -\frac{12r_0}{\zeta^4}\left[\left(\frac{2}{\eta} - \sin\eta - \frac{2}{\eta}\cos\eta\right)\zeta^2 + \left(1 + \cos\eta - \frac{2}{\eta}\sin\eta\right)(\zeta^2\coth\zeta - \zeta) + (\tfrac{1}{3}\zeta^3 - \zeta^2\coth\zeta + \zeta) + (\eta + \tfrac{1}{6}\eta^3 - 2\sin\eta + \eta\cos\eta)\right], \tag{60}$$

and the corresponding diagram is shown in Fig. 14. Here the curve A shows oscillation possibilities for transit angles as small as $3/2\pi$, while a much greater amount of resistance would have to be added to the circuit in order to eliminate the negative resistance and so stop the oscillations. In all, then, this method appears to be a better way of operating the system than with the positive plate, and this conclusion is substantiated by experimental observations.

As before, an increase in the grid capture fraction moves the oscillation region up to higher frequencies.

In both of the examples cited above, and represented by Figs. 11 and 14, respectively, complete space charge was assumed near the cathode. The effect of decreasing the cathode heating current so that this charge becomes negligible may be computed by employing (37) in place of (25), and (38) in place of (26).

The resulting equations for a slightly positive plate are,

$$R_0 = -\frac{12r_0}{\zeta^4}\left[\left(\frac{\eta-\sin\eta}{\eta}\right)(1-\cos\zeta)-\left(\frac{1-\cos\eta}{\eta}\right)(\zeta-\sin\zeta)\right], \quad (61)$$

$$X_0 = -\frac{12r_0}{\zeta^4}\left\{\left(\frac{\eta-\sin\eta}{\eta}\right)(\zeta-\sin\zeta)+\left(\frac{1-\cos\eta}{\eta}\right)(1-\cos\zeta)\right.$$
$$\left.+\left[\tfrac{1}{6}\zeta^3-\frac{4}{\zeta}(1-\cos\zeta)+2\sin\zeta\right]+\tfrac{1}{4}\eta\zeta^2\right\}. \quad (62)$$

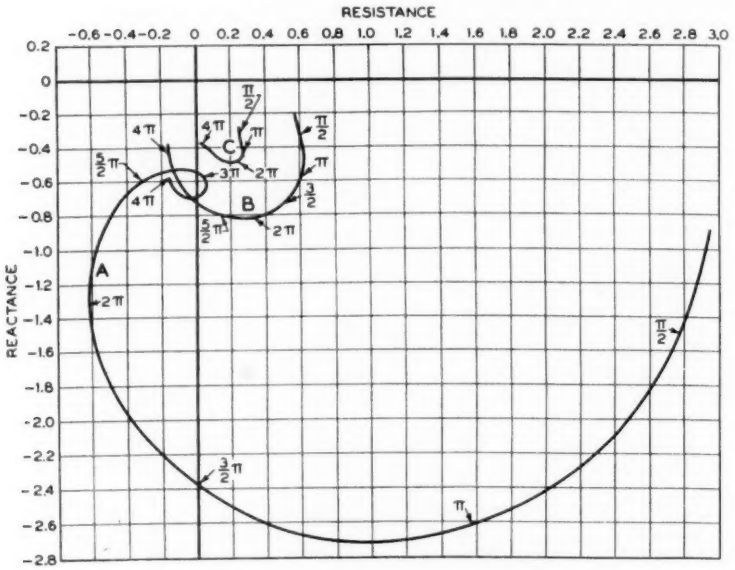


Fig. 14—$R_0 - X_0$ diagram for positive grid, slightly negative plate triode with cathode space charge.

Curve A, $\eta = \zeta$
Curve B, $\eta = 1/2\zeta$
Curve C, $\eta = 1/3\zeta$

The corresponding $R - X$ diagram is given in Fig. 15.

Again, the equations for a negative plate and no cathode space charge are,

$$R_0 = -\frac{12r_0}{\zeta^4}\left[\left(\frac{\eta-\sin\eta}{\eta}\right)\zeta^2-\left(\frac{1-\cos\eta}{\eta}\right)(\zeta^2\coth\zeta-\zeta)\right], \quad (63)$$

$$X_0 = -\frac{12r_0}{\zeta^4}\left[\left(\frac{1-\cos\eta}{\eta}\right)\zeta^2+\left(\frac{\eta-\sin\eta}{\eta}\right)(\zeta^2\coth\zeta-\zeta)\right.$$

$$+ \left(\tfrac{1}{3}\zeta^3 - \zeta^2 \coth \zeta + \zeta\right) + \tfrac{1}{4}\eta\zeta^2 \Big|, \tag{64}$$

and the $R - X$ diagram is shown in Fig. 16.

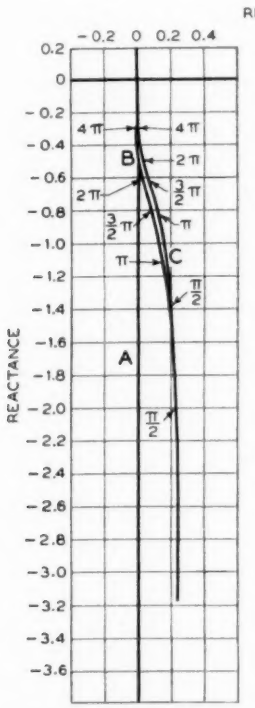


Fig. 15—$R_0 - X_0$ diagram for positive grid, slightly positive plate triode, without cathode space charge.

Curve A, $\eta = \zeta$
Curve B, $\eta = 1/2\zeta$
Curve C, $\eta = 1/3\zeta$

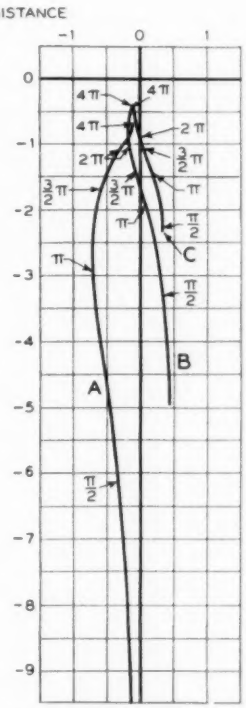


Fig. 16—$R_0 - X_0$ diagram for positive grid, slightly negative plate triode, without cathode space charge.

Curve A, $\eta = \zeta$
Curve B, $\eta = 1/2\zeta$
Curve C, $\eta = 1/3\zeta$

Inspection of Figs. 15 and 16 shows that the negative plate condition is greatly to be preferred when there is no cathode space charge. In fact, when account is taken of the difference in the scales for which Fig. 16 and the other three figures, 11, 14, and 15, are plotted, it is evident that the negative plate without space charge offers the greatest latitude in the adjustment of circuit condition. As in all of the cases, except Fig. 15, a small grid capture fraction is to be desired. If

curve A in Fig. 16 could be attained practically, it would be possible to secure oscillations even at low frequencies by connecting an inductance between the plate and cathode terminals. Curve B shows a low-frequency limit of a little less than $\frac{3}{2}\pi$ for the grid-plate transit angle.

One of the more important observations to be drawn from the curves of Figs. 11, 14, 15, and 16 is that if the inductance between plate and cathode is obtained by means of a tuned antiresonant circuit, then the circuit must be tuned to a frequency somewhat higher than the oscillation frequency. This is in order that it may effectively present an inductive impedance to the oscillating tube, so that the extended curves in the figures may encircle the origin in a clockwise direction.

Another conclusion is that there are so many different permutations and combinations of the operating conditions that it is small wonder that there have been a great many different "theories" and empirical frequency formulas advocated. For instance, operation under conditions giving an $R - X$ diagram which shows negative resistance over a small frequency range, only, such as A in Fig. 11, or B in Fig. 15, would give oscillations whose frequency would be much more nearly independent of the tuning of the external circuit than would conditions which resulted in a negative resistance over a wide frequency range, as at A in Fig. 16. In this latter case the external circuit exerts a large influence upon the frequency.

The data from which Figs. 11 to 16 were plotted are given in the appended tables. The final step in the calculation of these data was a multiplication by 12 which was performed on a slide rule. For all previous steps seven-place tables were employed because of the frequent occurrence of differences of numbers of comparable magnitude.

The effect on the frequency of a change in the operating voltages can be deduced inferentially from the curves. Thus, in general, the formulas for the transit angle have the form,

$$\xi = \frac{Kx}{\lambda\sqrt{V_0}}, \tag{65}$$

where x is the grid-origin distance,
λ is the wave-length,
V_0 is the grid potential,
K is a constant which depends on the mode of operation.

When the plate potential is changed by a relatively large amount the operation undergoes a transition from a limiting mode illustrated by one of the figures to another limiting mode shown on some other one of the figures.

On the other hand, a change in grid potential will act to change the

transit angles on the two sides of the grid in the same proportion. A modification of this generality occurs because the value of x in (65) will shift as the effective position of the virtual cathode moves about. Also, the complete space-charge condition near the plate becomes modified, and so the general relations become extremely variable. The partial space charge that exists with very negative values of plate potential, or with very high values of grid potential does not lend itself readily to mathematical treatment, so that intermediate conditions between complete and negligible space charge can be treated only by inference as to what happens between the two limiting conditions.

With inappreciable space charge on both the plate and the cathode sides of the grid, there can be no oscillations at all, since all impedances then approach pure capacities, with no negative resistance components.

A word concerning the so-called "dwarf" waves is in order before this general theoretical discussion is completed. In the curves, Fig. 14 distinctly shows this possibility in curve A, since the resistance reaches a large negative value at 2π and again at 4π. Likewise Fig. 11 shows the same possibility. On account of the resulting confusion in the figures, the higher frequency portions have not been drawn in the figures, but from (57), (59), (61), and (63) we can see what happens. Thus, for very high frequencies, η is large compared with unity, so that the formulas may be written,

$$R_0 = -\frac{12r_0}{\zeta^4}(\eta + \zeta)\sin\eta, \tag{57-a}$$

$$R_0 = -\frac{12r_0}{\zeta^2}(1 + \cos\eta + \sin\eta),$$

$$= -\frac{12r_0}{\zeta^2}\left[1 + \sqrt{2}\sin\left(\eta + \frac{\pi}{4}\right)\right], \tag{59-a}$$

$$R_0 = -\frac{12r_0}{\xi^4}\left[1 - \frac{\xi}{\eta}\cos\xi + \frac{\xi}{\eta}\cos\eta\right], \tag{61-a}$$

$$R_0 = -\frac{12r_0}{\xi^2}. \tag{63-a}$$

It is noteworthy that all of these exhibit the possibility of "dwarf" waves separated by discrete frequency intervals except (63-a). On the other hand, (63-a) gives possible conditions for operation at all high frequencies provided that the proper external circuit may be secured.

VIII. Postscript

The extension of the electronics of vacuum tubes which was described in the preceding pages must be regarded in the light of a tenta-

tive starting point rather than as a completed structure. Of the fundamental correctness of the method of attack there can be little doubt. The various simplifying assumptions, however, require careful scrutiny and doubtless some of them will be revised as time goes on and additional experience is acquired. Experimental guidance will be invaluable, and indeed certain data already have been obtained which are helpful in analysis of the assumptions. Although these data are in general qualitative agreement with the theory as outlined, the experimental technique must be refined before quantitative comparison can be made. It is hoped that the results can be made available at an early date.

Among the various assumptions which were made in the development of the theory, there are three which lead particularly to far-reaching consequences. These three may be enumerated as follows:

1. Plane-parallel tube structures
2. Current flow in straight lines
3. Small alternating-current amplitudes.

There are grounds for the belief that the assumption of plane-parallel tube structures does not exclude the application of the alternating-current results to cylindrical structures as completely as might be supposed. In the first place, the approximation of cylindrical arrangements to the plane-parallel structure becomes better as the cathode diameter is made large. Many tubes contain special cathode structures where this is the case. Furthermore, Benham [1] has obtained an approximate solution for the alternating-current velocity in cylindrical diodes where the cathode diameter is vanishingly small, and the transit angle is less than 5 radians. The resulting curves of alternating-current velocity versus transit angle have the same shape as the curves for the planar structures, and when the cylindrical transit angle is arbitrarily increased by about 20 per cent, the quantitative agreement is fair for transit angles less than 4 radians. It follows that until accurate solutions for cylindrical triodes can be obtained, the planar solutions may be expected to give correct qualitative results, and fair quantitative results when appropriate modifications of the transit angle are made. In fact, good agreement is obtained if calculations of the cylindrical transit angle are made as though the structure were planar.

The assumption of current flow in straight lines is open to some question when a grid mesh is interposed in the current path. For the positive grid triode, the objection to the assumption has been overcome by postulating a special case where an ideal choke coil prevents

the grid from carrying away any of the alternating current. Benham [1] has suggested an alternative which seems to work fairly well when the grid-cathode path of a negative grid tube is considered, but which offers grave difficulties when the grid-plate path is included. A still different alternative was employed in the present paper in connection with negative grid triodes, and successfully indicates phase angles for the mutual conductance of the tube which are qualitatively logical. The grid-plate path is still without adequate treatment, however.

As to the third general assumption: that of relatively small alternating-current amplitudes, there can be no objection from a strictly mathematical point of view, and for a very large proportion of the physical applications the assumption is thoroughly justified. Indeed, it is the only one which is successful in giving *starting* conditions for oscillators. However, when questions as to the power efficiency of oscillators or amplifiers arise, then the "small signal" theory is inadequate, and should be supplanted by an approximate theory. The form which this approximate theory should take is indicated by the standard methods of dealing with the efficiencies of low-frequency power amplifiers and oscillators where the wave shape of the plate current is assumed to be given. The application of the same kind of approximation to ultra-high-frequency circuits may eventually prove to be a simpler matter than the "small signal" theory set forth in these pages.

Besides the three main assumptions discussed above, there was a fourth assumption which, although of lesser importance, deserves some comment. This fourth assumption involves the neglect of initial velocities at a hot cathode. If all electrons were emitted with the same velocity, the theory is adequate, and may be applied as indicated by Langmuir and Compton.[3] When the distribution of velocities according to Maxwellian, or Fermi-Dirac, laws is considered, some modifications may be necessary. In general, a kind of blurring of the clear-cut results of the univelocity theory may be expected, which will be expected to result in an increase in the resistive components of the various impedances at the expense of the reactive components. Again, lack of symmetry in the geometry of the tube structure may be expected to do the same thing, since the transit angles are then different in the different directions.

Finally, however, and with all its encumbering assumptions, it is hoped that the excursion back to fundamentals which was made in this paper, has resulted in a method of visualizing the motions of the condensations and rarefactions of the electron densities inside of vacuum tubes operating at high frequencies and has shown their relation to the conduction and displacement components of the total current.

DATA FOR FIG. 11

ξ	$\eta = \xi$		$\eta = \frac{1}{2}\xi$		$\eta = \frac{1}{3}\xi$		$\eta = \frac{1}{4}\xi$	
	R_0	X_0	R_0	X_0	R_0	X_0	R_0	X_0
1.0	1.87	−0.577	0.302	−0.119			0.0633	
1.4	1.75	−0.777	0.292	−0.164			0.0604	−0.160
1.57	1.69	−0.855	0.288	−0.182	0.112	−0.172		
1.8	1.60	−0.949	0.280	−0.204	0.108	−0.193	0.0568	−0.200
2.356	1.36	−1.13	0.260	−0.241	0.0987	−0.240		
2.8	1.15	−1.23	0.241	−0.289			0.0458	−0.278
3.14	0.986	−1.27	0.226	−0.311	0.0838	−0.289	0.0418	−0.297
3.6	0.769	−1.29	0.204	−0.335	0.0748	−0.308	0.0364	−0.316
4.0	0.593	−1.27	0.186	−0.350	0.0673	−0.320	0.0320	−0.327
4.71	0.326	−1.18	0.153	−0.366	0.0457	−0.329		
5.2	0.186	−1.08	0.132	−0.368			0.0216	−0.330
5.6	0.0972	−0.987	0.116	−0.368			0.0193	−0.324
6.28	0	−0.830	0.0923	−0.358	0.0365	−0.309	0.0165	−0.306
6.8	−0.0348	−0.719	0.0767	−0.346			0.0153	−0.291
7.2	−0.0440	−0.644	0.0662	−0.337	0.0308	−0.284	0.0148	−0.278
7.85	−0.0369	−0.546	0.0515	−0.318	0.0284	−0.264		
8.4	−0.0199	−0.487	0.0414	−0.302	0.0268	−0.248	0.0144	−0.238
9.0	+0.000414	−0.444	0.0320	−0.284	0.0254	−0.233	0.0144	−0.222
9.42	0.0125	−0.424	0.0262	−0.272	0.0244	−0.222	0.0143	−0.235
10.0	0.0218	−0.407	0.0194	−0.256			0.0138	−0.198
10.99	0.0213	−0.385	0.0101	−0.232	0.0192	−0.193		
12.57	0	−0.342	0	−0.197	0.0128	−0.172	0.00962	−0.162

DATA FOR FIG. 14

ξ	$\eta = \xi$		$\eta = \frac{1}{2}\xi$		$\eta = \frac{1}{3}\xi$		$\eta = \frac{1}{4}\xi$	
	R_0	X_0	R_0	X_0	R_0	X_0	R_0	X_0
1.0	2.94	−0.932	0.581	−0.225			0.133	−0.221
1.4	2.84	−1.33	0.595	−0.304			0.136	−0.286
1.57	2.78	−1.49	0.601	−0.336	0.252	−0.302		
1.8	2.67	−1.71	0.606	−0.377	0.255	−0.330	0.139	−0.336
2.356	2.30	−2.19	0.618	−0.465	0.264	−0.381		
2.8	1.90	−2.48	0.619	−0.528			0.147	−0.401
3.14	1.56	−2.63	0.613	−0.571	0.272	−0.424	0.150	−0.410
3.6	1.06	−2.71	0.598	−0.625	0.276	−0.438	0.153	−0.416
4.0	0.645	−2.67	0.577	−0.665	0.277	−0.448	0.155	−0.417
4.71	0.00015	−2.38	0.525	−0.728	0.275	−0.460	0.157	−0.412
5.2	−0.319	−2.08	0.479	−0.762			0.158	−0.407
5.6	−0.494	−1.80	0.436	−0.784			0.158	−0.404
6.28	−0.608	−1.31	0.356	−0.807	0.253	−0.476	0.156	−0.396
6.8	−0.565	−0.990	0.292	−0.814			0.154	−0.390
7.2	−0.480	−0.796	0.241	−0.803	0.232	−0.483	0.152	−0.386
7.85	−0.290	−0.594	0.160	−0.797	0.213	−0.485		
8.4	−0.111	−0.528	0.0957	−0.773	0.196	−0.486	0.144	−0.376
9.0	+0.00226	−0.536	0.0308	−0.735	0.175	−0.485	0.138	−0.372
9.42	0.0574	−0.574	−0.0102	−0.705	0.160	−0.484	0.133	−0.369
10.0	0.077	−0.634	−0.0581	−0.666			0.127	−0.365
10.99	0	−0.690	−0.118	−0.564	0.101	−0.436		
12.57	−0.152	−0.560	−0.152	−0.414	0.0445	−0.383	0.0938	−0.348

DATA FOR FIG. 15

ξ	$\eta = \xi$		$\eta = \frac{1}{2}\xi$		$\eta = \frac{1}{3}\xi$		$\eta = \frac{1}{4}\xi$	
	R_0	X_0	R_0	X_0	R_0	X_0	R_0	X_0
1.0			0.239	−3.06			0.179	−1.58
1.4			0.228	−2.22			0.172	−1.19
1.57			0.223	−2.00	0.199	−1.39		
1.8			0.215	−1.76	0.192	−1.24	0.162	−0.979
2.356			0.193	−1.39	0.173	−1.01		
2.8			0.173	−1.21			0.131	−0.737
3.14			0.157	−1.10	0.130	−0.825	0.120	−0.693
3.6			0.135	−0.983	0.123	−0.756	0.104	−0.645
4.0			0.116	−0.903			0.0902	−0.610
4.71			0.0837	−0.788	0.0796	−0.633		
5.2			0.0643	−0.724			0.0540	−0.524
5.6			0.0503	−0.677				
6.28			0.0308	−0.605	0.0346	−0.506	0.0308	−0.455
6.8			0.0197	−0.558			0.0232	−0.425
7.2			0.0131	−0.525	0.0194	−0.444	0.0187	−0.402
7.85			0.00568	−0.477	0.0126	−0.406		
8.4			0.00203	−0.442	0.00939	−0.378	0.109	−0.343
9.0			−0.0000284	−0.409	0.00710	−0.351	0.00908	−0.318
9.42			−0.000646	−0.388	0.00608	−0.333	0.00826	−0.290
10.0			−0.000817	−0.363			0.00744	−0.284
10.99			−0.000402	−0.327	0.00408	−0.282		
12.57			0	−0.285	0.00216	−0.246	0.00385	−0.227

DATA FOR FIG. 16

ξ	$\eta = \xi$		$\eta = \frac{1}{2}\xi$		$\eta = \frac{1}{3}\xi$		$\eta = \frac{1}{4}\xi$	
	R_0	X_0	R_0	X_0	R_0	X_0	R_0	X_0
1.0	−0.176	−9.35	0.426	−4.84			0.342	−2.52
1.4	−0.306	−6.84	0.366	−3.64			0.316	−1.96
1.57	−0.363	−6.15	0.338	−3.33	0.345	−2.32		
1.8	−0.436	−5.41	0.299	−3.00	0.321	−2.11	0.287	−1.67
2.356	−0.584	−4.15	0.206	−2.46	0.263	−1.77		
2.8	−0.658	−3.44	0.137	−2.17			0.212	−1.29
3.14	−0.685	−3.00	0.0888	−1.99	0.188	−1.48	0.190	−1.21
3.6	−0.685	−2.52	0.0318	−1.80	0.149	−1.36	0.162	−1.12
4.0	−0.661	−2.17	−0.0104	−1.65	0.120	−1.27	0.140	−1.06
4.71	−0.565	−1.69	−0.0698	−1.43	0.0747	−1.13	0.108	−0.957
5.2	−0.482	−1.45	−0.0998	−1.30			0.0871	−0.898
5.6	−0.413	−1.30	−0.119	−1.21			0.0730	−0.853
6.28	−0.304	−1.11	−0.141	−1.07	0.0478	−0.907	0.0523	−0.787
6.8	−0.236	−1.02	−0.151	−0.975			0.0389	−0.742
7.2	−0.195	−0.963	−0.155	−0.910	−0.0220	−0.805	0.0296	−0.709
7.85	−0.148	−0.894	−0.156	−0.815	−0.0364	−0.743		
8.4	−0.126	−0.848	−0.152	−0.747	−0.0458	−0.695	+0.0072	−0.625
9.0	−0.113	−0.803	−0.145	−0.680	−0.0538	−0.647	−0.00162	−0.589
9.42	−0.109	−0.771	−0.138	−0.638	−0.0582	−0.615	−0.00706	−0.565
10.0	−0.107	−0.727	−0.128	−0.588			−0.0135	−0.535
10.99	−0.101	−0.653	−0.107	−0.518	−0.0668	−0.517		
12.57	−0.0760	−0.556	−0.0760	−0.439	−0.0667	−0.439	−0.0314	−0.426

Contemporary Advances in Physics, XXVII
The Nucleus, Second Part *

By KARL K. DARROW

In this Second Part the major subject is Transmutation: that is to say, the alteration or disintegration of a nucleus, the unique and distinctive part of any atom, by impacts of fast-moving corpuscles. For the last year and a half the pace of progress in this field has been increasingly rapid, and in all likelihood is destined to become yet swifter. This is partly because of the discovery—a discovery due largely to theoretical foresight—that transmutation of some elements is practicable with protons of a relatively modest energy which can be produced in laboratories without any serious difficulty. Partly it is due to the discovery of neutrons and of deutons, particles which apparently possess remarkable ability in effecting certain kinds of transmutation. Partly also it is due to advances and refinements in the methods of working with alpha-particles, the first variety of corpuscle with which disintegration of nuclei was ever achieved. People are already beginning to speak of "nuclear chemistry" as a special branch of science, and this is already almost justified by the number of cases known in which two nuclei interact and produce two others which are recognizable.

Bringing the First Part up to Date

STRANGE as it seems to speak of "bringing up to date" something that was published only six months ago, one is sometimes obliged to do so by the rapid march of science; and three of the "elementary particles" of which I spoke in the First Part were and still are so young—or to speak more carefully, our acquaintance with them is still so young—that their rôle and situation in the body of physical knowledge is changing from month to month.

The Positive Electron

Of the positive electron the most striking new thing to be said is, that there is now a new way of generating it: by impacts of alpha-particles against metals. This so far has been applied only by its discoverers, M. and Mme. Joliot; only with alpha-particles from polonium, therefore of energy 5.3 millions of electron-volts; only to five metals, of which beryllium and boron and aluminium yielded positive electrons, while silver and lithium did not. It is as yet the most efficacious way of producing positive electrons, Joliot having evoked last summer as many as 30,000 of these corpuscles per second from aluminium. This of course looks small when compared with the torrents of negative electrons which incandescent metals will pour out,

* "The Nucleus, First Part" was published in the July 1933 issue of the *Bell Sys. Tech. Jour.*, Vol. XII.

but these are not a proper standard of comparison. Rather should one say that in the autumn of 1932 positive electrons were being observed at the rate of three or four a year, and already by the summer of 1933 this rate had been enhanced to thirty thousand in the second!

The other voluntary way of generating positive electrons—by applying hard gamma-rays to heavy elements—has already been studied enough to yield the data of the following table. Here, in the first column, stand the names of various sources of gamma-rays (the one denoted as "Po + Be" is beryllium exposed to impacts of alpha-particles from polonium); in the second, the energy-values in MEV (I use this symbol hereafter for "millions of electron-volts") of the individual photons of these rays; in the third, the symbols of various metals; in the fourth, the number of positive electrons per hundred negatives, ejected from these metals by these gamma-rays; in the fifth, the authorities:

Po + Be	5	U	40	Joliots
		Pb	30	Joliots
		Pb	35	Chadwick
		Cu	18	Joliots
		Al	5	Joliots
ThC"	2.6	Pb	8	Joliots
		Pb	4	Chadwick
Ra(B + C)	1.0–2.2	Pb	3	Grinberg
Po	0.85	Pb	0	Meitner-Philipp

The percentages in the fourth column give at the moment our best available notion as to the relative plentifulness of positive electrons, produced by the several kinds of rays falling upon the several metals. One would prefer to have the total number of positives per unit intensity of the infalling rays, but that is not available at present—I presume because of the difficulty of measuring these intensities. One must remember that the data usually consist in observations of a few hundred or a few dozen cloud-tracks, so that the accuracy of these percentages cannot be great.[1]

We note that with lead the proportion of positive electrons mounts rapidly with increasing photon-energy, and that with 5 MEV-photons

[1] This perhaps is sufficient to account for a discrepancy between the general trend of the table and a value of 1/3 given by Meitner and Philipp for the ratio of positives to negatives when brass is exposed to (Po + Be). Should the table be extended and supported by a successful theory, it should then be possible to determine the frequency of gamma-rays by the percentage of positives which they produce when falling on a metal. In this connection it is interesting that Anderson's latest data indicate that positive and negative electrons are about equally abundant among the ionizing particles of the cosmic rays, a fact which suggests that if they are due to photons, these must be of a distinctly higher energy than any of those cited in the foregoing table.

the proportion goes up rapidly with the nuclear mass of the bombarded atoms.[2] Both of these rules are in harmony with the remarkable theory to which I alluded in the First Part—the theory that each positive electron (together with a negative companion) springs into being from a transmutation of light into electricity! It is supposed that a photon transmutes itself into a pair of electrons, one of each sign.

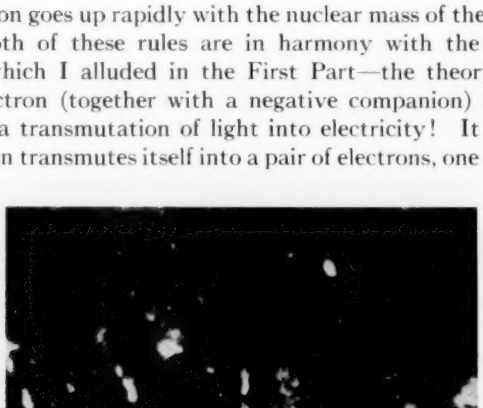


Fig. 1—Tracks of an electron-pair (positive and negative) arising in argon exposed to gamma-rays, and probably created near an argon nucleus by transmutation of a photon. (M. et Mme. Joliot)

Conservation of the net charge of the universe is assured in this hypothetical process. Conservation of mass and energy is attainable, for the speeds of the electrons may be such that their energies together are equal to the energy of the vanished photon. I take this occasion to repeat Einstein's principle, which figures so importantly in these articles. The energy E of a material particle moving with speed v (relatively to the observer) is given by the formula:

$$E = m_0(1 - \beta^2)^{-1/2}c^2 = mc^2,$$

[2] In this connection it should be noted that the source "Po + Be" emits neutrons as well as photons, and while the first-named are certainly not chiefly responsible for the positive electrons, they may produce some of these. Chadwick observed that the rays from a "Po + B" source (boron in place of beryllium), which consist of neutrons plus some photons of about the same energy as the photons of ThC", evoked from lead a distinctly larger percentage of positive electrons than do the rays of ThC".

in which c stands for the speed of light in vacuo; β for v/c; and m_0 for a constant. The ratio of E to c^2 is the function of v and m_0 which this equation defines, and is denoted by m and called the mass of the particle: it is in this sense that mass and energy are equivalent. We may (mentally) divide the mass of the moving particle into two terms m_0 and $(m - m_0)$, and the energy into two terms m_0c^2 and $(m - m_0)c^2$. We may further call m_0 the rest-mass and m_0c^2 the energy associated with the rest-mass; and we may call $(m - m_0)c^2$ the kinetic energy and $(m - m_0)$ the mass associated with the kinetic energy or the extra mass due to the motion of the particle. Such will be the terminology used in these articles, although this definition of kinetic energy is only approximately the same as the classical and familiar one.[3]

Returning to the argument about the transmutation of light into electrons, or more precisely, of a photon into an electron-pair: conservation of mass and energy is attainable, for the two electrons may have such speeds—call them $\beta_1 c$ and $\beta_2 c$—that the sum of $m_0(1 - \beta_1^2)^{-1/2}c^2$ and $m_0(1 - \beta_2^2)^{-1/2}c^2$ is equal to the energy $h\nu$ of the photon. But the demand for conservation of momentum makes apparently serious trouble. If we assume that a photon voyaging through the depths of space suddenly converts itself spontaneously into a pair of electrons, and if then we attempt to impose both conservation of momentum and conservation of energy, the equations lead us straightway into an inescapable muddle, in which the original assumptions contradict each other. We are driven therefore to infer that the imagined process is impossible. But this seeming catastrophe of the theory turns out to be a blessing. What is observed is not after all the transmutation of a photon in the depths of empty space, but a process which occurs in the depths of plates of lead and other heavy elements. If we suppose that such a transmutation occurs near to a massive nucleus, then this may receive some of the energy and some of the momentum of the photon; and the equations show that the momentum which it takes may be quite sufficient to permit the process to occur, while the energy which it takes is so small that for practical purposes we may still pretend that the whole of the energy of the photon is divided between the electrons (though we certainly should not forget about the small fraction which goes to the nucleus). All the principles are thus fulfillable: conservation of charge requires that there should be

[3] The classical definition of kinetic energy is $(1/2)mv^2$; the present or relativistic definition, viz. $(m - m_0)c^2$, is an infinite series of which the first term is identical with the classical definition. The difference between the two definitions increases as the speed of the particle increases, but so far as I know there has not yet been an actual case in which it is of practical importance.

two electrons of opposite signs; conservation of energy, that they should have appropriate speeds; conservation of momentum, that the process should occur only near a massive nucleus.

This is the most alluring of all theories, for it is the doctrine that the substance of matter and the substance of light are ultimately the same, being interconvertible. It therefore demands, and is surely destined to receive, the sharpest and fullest of testing; the more so because there is a rival in the field, the theory that the positive electron exists beforehand and from all time in the nucleus of the atom, and is ejected from it by the photon. The newest way of producing positive electrons by alpha-particle impact seems to speak in favor of the latter. One could indeed suppose that the kinetic energy of the alpha-particle is transformed into an electron-pair, directly or through the intermediacy of a transiently-existing photon; but this would be an artificial idea unless it were to be supported by a basic theory or by observing that the positive electrons are often paired with negatives (which the Joliots do not say). In favor of the former theory speak the facts that in several scores of cases paired electrons have been observed—*i.e.*, two electrons of opposite signs were seen to spring from the same point (so far as the eye could tell)—when metal plates were bombarded with gamma-rays; and the further fact that the energies of these electron-pairs and of individual positives did not surpass those of the infalling photons, though they approached it often.[4] There are always apparently unpaired positives and many more unpaired negatives; but one may always say that with some of the pairs it happened that one member remained in the metal and the other got away, while many of the negatives are surely electrons which have been expelled from their places by photons acting in the well-known ways. Further, there are more or less forcible indications that some part of the absorption of gamma-rays in heavy metals may be ascribed to the formation of electron-pairs, and some part of the radiation scattered from the metals when gamma-rays fall on them may be attributed to the reunion of two electrons of opposite sign which re-transmute themselves into light; but some of the data are not checked, and the time seems not ripe for reviewing them. In the hands of Oppenheimer and Plesset the transmutation theory has supplied other quantitative

[4] See First Part of this article, pp. 304–305, *B.S.T.J.*, July 1933. The *kinetic* energies of electron-pairs and *a fortiori* of positive electrons should not come within one million electron-volts of the energy-value of the photons, for the rest-mass of two electrons amounts approximately to a million of these units. This rule has lately been strengthened by evidence from Anderson and his colleagues, who in a couple of hundred of additional cases find no violation of it; the distribution-in-energy curves for pairs and for (apparently) isolated positives extend up to the predicted upper limit, and there they fall to the horizontal axis. More evidence of this kind has been accumulated by Blackett (loc. cit. footnote 5).

predictions meet for testing, and it is likely that in six months more a great deal will be learned.[5]

The Deuton

The newly-discovered isotope of hydrogen of mass-number 2—H^2, "heavy hydrogen," or, to adopt Urey's name for it, "deuterium"—has suddenly become the most popular and the most eagerly sought-after of all chemical substances. This is because of the notable chemical and physical differences between it and its compounds on the one hand, H^1 and the corresponding compounds of H^1 on the other. So great are these differences that by the usage of twenty years ago H^2 would probably have been called a new element, and indeed it deserves all the prestige that would accrue to it from being so denoted; but to violate the present and most wisely-based of usages, whereby an element is characterized by atomic number rather than by the ensemble of its properties, would be mistaken.[6]

Deuterium is so rare by comparison with H^1 (Urey's "protium") that it would still be very unfamiliar, but for the unexpected and remarkable efficacy of the electrolytic method of separating water molecules comprising H^2 atoms from water molecules comprising none but H^1 atoms. It turns out that if an aqueous solution is electrolyzed until only a very tiny fraction of the original liquid remains, the proportion of the former kind of molecule in that tiny residue is anomalously large. Washburn seems to have been the first to suspect that this might happen; he procured samples of the residues from electrolytic cells which had been operated continuously in commercial plants for two and three years, and sent them to Urey, who performed a spectrum-analysis and observed "a very definite increase in the abundance of H^2 relative to H^1." Shortly afterwards the method was put into operation on a grand scale by G. N. Lewis and his collaborators, with spectacular results. In one experiment, for instance, they started out with twenty liters of water, electrolyzed it until there remained but half a cc. of liquid, and found that in this residue deuterium atoms made up two-thirds of all the hydrogen atoms which were left. For months thereafter, nearly every paper on deuterium and on the deuton which was published began with an acknowledgment to Lewis for a small amount of water rich in heavy hydrogen which the fortunate author had received from him.

[5] For a fuller account of the situation as it now stands, see an article of mine in the *Scientific Monthly*, January 1934; also one by P. M. S. Blackett, *Nature* **132**, pp. 917–919 (Dec. 16, 1933), which incidentally contains some further data.

[6] I should think that the case of deuterium by itself would make it necessary henceforth to define the concept "element" altogether from the concept "atomic number," forsaking all the earlier definitions.

Interesting as are the chemical and physical properties of deuterium and its compounds, we are here concerned only with the nucleus of the H^2 atom, the deuton (all the other suggested names seem to be fading out). The accepted value for its mass is that given by Bainbridge, 2.0131 on the standard scale in which the mass of the O^{16} atom is 16 exactly. Of its spin I shall speak in a later article. Its powers of transmutation are remarkable, and quite unlike those of H^1; if first a beam of H^1 nuclei (protons) and then an equal beam of deutons be directed against targets of various elements, the number of fragments observed per unit time is greater for some elements and less for others, and their ranges in general are different. In some cases it seems possible that the deutons themselves are being split into protons and neutrons, a result of great importance if it can be established beyond question. We shall consider the data at length.

The neutron

Most of what has newly been learned about the neutron will find appropriate places elsewhere in this article. There should be a separate section about the deflections suffered by neutrons when they impinge on or pass close to nuclei without transmuting them—the topic known as "scattering," "interception," or (badly) "absorption" of neutrons. This topic however is scarcely ripe for description in such an article as this, the experiments being difficult and the inferences from the data being highly controversial. I therefore postpone it to some future occasion, remarking only that it seems established that a neutron may pass within a very short distance indeed from a nucleus—only a very few times 10^{-13} cm from the centre thereof—without interacting with that nucleus in any perceptible way.

Masses of the Lighter Atoms

There are now thirteen of the lighter atoms of which the masses—in terms of the mass of the O^{16} atom taken as 16 exactly—have been determined to four and even to five significant figures. Most of these values were mentioned in the First Part, but it will be convenient to have them all tabulated here. They are the masses of complete atoms, nuclei accompanied by their full quotas of orbital electrons. The uncertainties quoted are the "probable errors"; where Aston originally gave the maximum possible uncertainty, this has been divided by 3 (see First Part, footnote 10). Values marked with an asterisk are from Bainbridge, the others from Aston; the value for H^1 has been obtained by both.

H^1	1.007775 ± .000035	C^{12}	12.0036 ± .0004
* H^2	2.01363 ± .00008	N^{14}	14.008 ± .001
He^4	4.00216 ± .00013	O^{16}	16.0000 (standard)
* Li^6	6.0145 ± .0003	F^{19}	19.000 ± .002
* Li^7	7.0146 ± .0006	* Ne^{20}	19.9967 ± .0009
* Be^9	9.0155 ± .0006	* Ne^{22}	21.99473 ± .00088
B^{10}	10.0135 ± .0005	* Cl^{35}	34.9796 ± .0012
B^{11}	11.0110 ± .0005	* Cl^{37}	36.9777 ± .0019

The table of the chemical atomic weights reproduced in the First Part has suffered two alterations: a very slight change in the given value for K, from 39.10 to 39.096; and an important change in the chemical atomic weight of carbon, which rises from 12.00 to 12.011, and now permits of an abundance of C^{13} easier to reconcile with the observed intensities of the spectrum-lines of this substance than was the abundance, or rather the scarcity, implied by the former value.[7]

The list of isotopes detected by Aston's mass-spectrograph has been enlarged by the following examples,[8] which the reader may enter upon Fig. 6 of Part I: neodymium, $Z = 60$, $A-Z = 83$; samarium, $Z = 62$, $A-Z = 85, 86, 87, 90, 92$; europium, $Z = 63$, $A-Z = 88, 90$; gadolinium, $Z = 64$, $A-Z = 91, 92, 93, 94, 96$; terbium, $Z = 65$, $A-Z = 94$.

New Developments in Transmutation: The Apparatus

In the two years and a quarter which are all that have elapsed since I published in this *Journal* an article on transmutation,[9] the situation in this field has vastly changed, and the prospects for the future have been amplified immensely. So lately as the early spring of 1932, disintegration of a nucleus had not yet been demonstrably achieved except by alpha-particles possessing energy not smaller than three millions of electron-volts. Schemes for producing five- and ten-million-volt ions were already under way, being ardently pushed onward because it was supposed that transmutation would never be effected by any agency much feebler. But in the course of 1930, Cockcroft and Walton of the Cavendish Laboratory had been emboldened by a theory (I will describe it later) to imagine that protons of only a few hundred thousand electron-volts might be able to transmute, and to risk their time and labors in the task of developing powerful streams of such particles. After two years of work they

[7] See an item in *Nature*, **132**, 790–791 (Nov. 18, 1933). In the table of masses on p. 303 of the First Part, change 1.0078 to 1.0072 and 4.002 to 4.001 (the former values refer to complete atoms, not bare nuclei).

[8] F. W. Aston, *Nature* **132**, 930–931 (Dec. 16, 1933).

[9] "Contemporary Advances in Physics XXII," *Bell System Technical Journal*, **10**, 628–665 (October 1931). I refer to this article hereinafter as *Transmutation*.

were justified in the event; for they detected fragments proceeding from targets of lithium bombarded by their protons, with energy-values anywhere from half-a-million down to only seventy thousand electron-volts.

It would be hard to overstate the joyful surprise of this announcement. Transmutation, of some elements at least, was easier by far than had been thought! It would not after all be necessary to fare forth into the unknown, and face at once the problems of applying voltages without precedent; successes which had seemed doubtful at best and assuredly distant were after all to be had by a relatively slight extension of a known technique. All over Europe and America people began making plans for applying these voltages, so much less formidable than those which had previously been thought indispensable. Nevertheless the first who confirmed and extended the work of Cockcroft and Walton were those who had aimed from the start at the higher and harder goal: Lawrence and his colleagues at Berkeley. Their work had not been wasted, for they instantly found themselves able to measure the disintegration of lithium by protons all the way up to 710,000 electron-volts; and within four months they had carried the upper limit onward to 1,125,000, and as I write these lines they have just announced that the limit has soared to three millions! From Pasadena also comes word of transmutation achieved by protons, and deutons, and helium nuclei, endowed with energy by voltages ranging downward from nine hundred thousands.

These are not the only novel results of the last two years and a quarter. The neutron has disclosed itself not only as a product, but as an agent of transmutation, able to alter nuclei which have thus far resisted both the alpha-particles and the protons which have been showered upon them in laboratories. The disintegrations effected by alpha-particles have been studied with ever-increasing minuteness and detail, and are beginning to show that nuclei are structures capable of existing in various normal states and excited states, characterized by distinctive energy-values. The emission of alpha-particles from radioactive nuclei has been studied with a new precision, and leads to the same conclusion. The astonishing feats achieved with bombarding particles of lesser energy have not lessened the hope of achieving startling things with particles of greater.

Cockcroft and Walton, inspired by theory, had built an apparatus for producing half-million-volt protons, and had proved them able to transmute. The proton-streams had not, however, been greater than five microamperes (one microampere or $\mu a = 6.28 \cdot 10^{12}$ protons per second). Next Oliphant and Rutherford, inspired by that result,

proceeded to build an apparatus in which the maximum voltage should not go above a quarter of a million, but in recompense the stream of protons should be raised to a hundred microamperes. Another alteration: previously the stream had been a mixture of protons with heavier ions and neutral particles—now Oliphant and Rutherford introduced a magnetic field, adjustable and strong enough to bring either the protons or the more massive ions separately against the target. The magnetic field also assures that all the particles striking the target shall have nearly the same speed, something not completely guaranteed by the constancy of the voltage.

The scheme of this device is sketched in Fig. 2, where the course of the proton-stream is traced (rather too pictorially, I fear!) in a sweeping arc from its origin in the discharge-tube *R*, to the target *T* where the element to be transmuted awaits the impacts. In the discharge-tube all the parts are of steel, and the block *C* and cylinder *B* conjointly form the cathode, while the oil-cooled block *D* and cylinder *A* conjointly form the anode. This unusual material and structure are required partly to minimize cathode-sputtering, and partly to take care of the great amount of heat which is steadily developed in the tube, inasmuch as for the best supply of protons a voltage of 20,000 and a current of many milliamperes are demanded. Something like a twentieth of the current in the discharge is borne through the hole in the cathode by protons (and other positive ions of greater mass, if such there be); and in the space between *C* and *E* these particles receive from an electric field most of the kinetic energy with which they strike the target. In this space and in the region where the magnetic field comes into play, the density of the gas must be kept extremely low, despite the fact that there is an open passage into these spaces from the discharge-tube where the density must always be great enough to sustain the discharge and the supply of protons. This is a task for powerful pumps, which must be kept continuously at work pumping away from the lower chambers the gas which is steadily draining out of the discharge-tube through the hole and must as steadily be replenished by feeding fresh hydrogen in from above. It is no small part of the difficulty of the experiment, that the discharge-tube and the source of its power and the source of its hydrogen must all be maintained at scores or hundreds of thousands of volts above the potential of the ground, in order that the observing-apparatus may itself be at ground-potential. The transmutations are observed by detecting the fragments which issue through the very thin mica pane of the window *W*.

Until the building of this apparatus proton streams had been so scanty, that to bring about disintegrations in measurable number it had been needful to project the protons against thick layers of dense matter. In going through these layers they were slowed down and

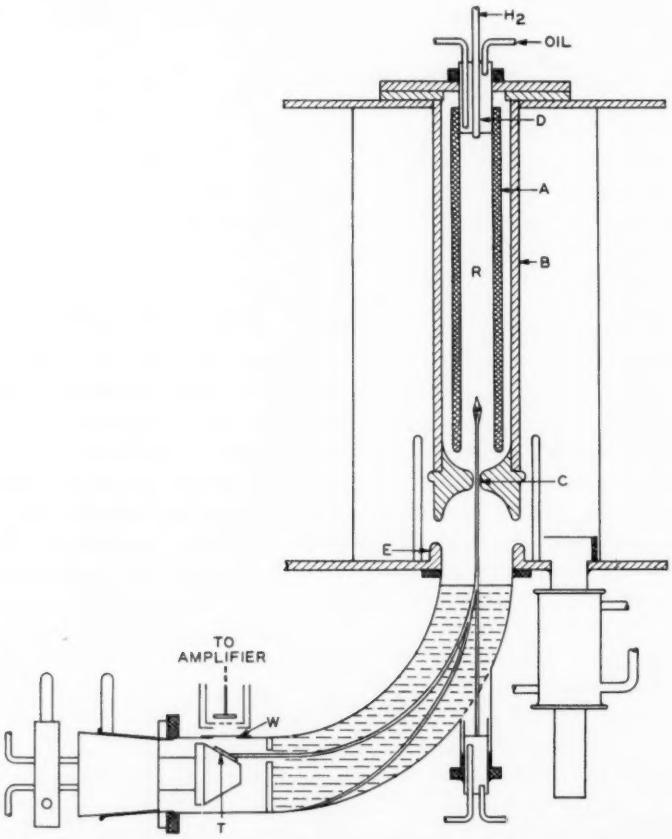


Fig. 2—Apparatus of Oliphant and Rutherford for producing transmutation by intense streams of protons. (*Proceedings of the Royal Society*)

stopped, and there was no direct way of telling whether the observed transmutations were achieved by protons of full speed, or by those which had already lost some energy, or both. Moreover if the energy of the bombarding particles was raised, the number of disintegrations infallibly went up, but a part of this increase (and sometimes the

whole of it) was certainly due to the fact that the faster particles went farther into the layer and struck more nuclei. There is no need to labor the point: it is obviously desirable to do the experiments with a film so thin that each oncoming proton either strikes a nucleus with its full and unabated initial energy, or else goes through the film and away without any impact at all. This ideal was closely approached by Oliphant and Rutherford, when they got countable numbers of fragments from films of lithium and boron (deposited on blocks of steel or iron) which were so thin as to be invisible, and of which the latter was known to consist of only seven-tenths as many atoms as would suffice to cover the iron surface with a single monatomic layer. (The curves of Fig. 16 were obtained with these films.)

This is a success which proves it possible to investigate films consisting each of only a single isotope of the element in question; for feeble as are the ways of separating isotopes in all but a few very favorable cases, they yet are powerful enough to produce pure monatomic layers. This article will amply show how valuable will be the privilege of getting data from a single isotope, of lithium or boron for example; already there are several cases of important antagonistic theories, the decisions between which will be given once and for all by such data.

The apparatus devised by E. O. Lawrence and developed in his school at Berkeley is of a singular ingenuity, inasmuch as in it ions are accelerated until their energies are such as would be derived from an unimpeded fall through a potential-difference of literally millions of volts, and yet the greatest voltage-difference at any moment between any two points of the apparatus is only a few thousands. It owes its elegant compactness to the lucky fact that when a charged particle is moving in a plane at right angles to a constant magnetic field, and consequently is describing a succession of circles, the time which it takes to describe a single circle is the same whatever its speed. One sees this readily by writing down the familiar equation,

$$mv^2/\rho = Hev/c,$$

in which e, m, v stand for the charge (in electrostatic units), mass, and speed of the ion and ρ for the radius of curvature of the circle, and on the right we have the force exerted by the magnetic field H upon the ion and on the left the so-called "centrifugal force" to which it is equal. The radius ρ varies directly as v, but the time $T = 2\pi\rho/v$ which the ion takes to describe a circle is independent of v. This is no longer true if the ion is moving so fast that the foregoing classical equation must be replaced by its relativistic analogue, but fortunately

the desired results are attained without forcing the speed to such heights.

Suppose now that while the ion is describing its consecutive circles each in a time T, its speed is suddenly increased; it continues to make circles, of a larger radius but with the same duration. Suppose that the increase occurs twice in each cycle, at intervals $T/2$; the path is a succession of semicircles each broader than the one preceding but all described in equal time. Now we arrive at Lawrence's device. The ions circulate in a round flat metal box, sliced in two along one of its diameters (Figs. 3, 4); and every time that one of them passes from

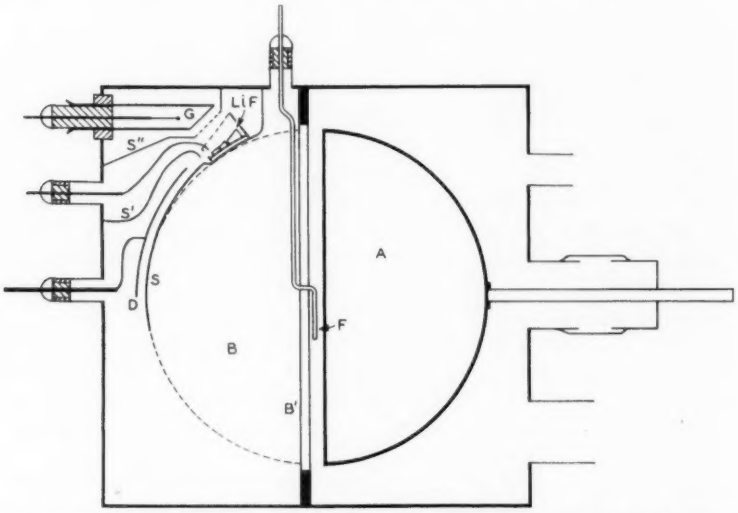


Fig. 3—Diagram of the Lawrence apparatus for cumulative accelerations of protons and other ions with auxiliary magnetic field. (After Henderson)

within half-box B to within half-box A it is accelerated by a voltage-difference existing between B and A, and every time that it passes from within A to within B it is again accelerated by a voltage-difference between B and A. Of course if this voltage-difference remained the same, the ion would lose at the latter passage just the energy which it gained at the former; but here is precisely the distinctive feature of the method: *the potential-difference between the two half-boxes is reversed in sign between each two consecutive passages.* So rapidly do the successive passages follow on one another, that if the intervals between them were unequal it would probably be impossible to devise any mechanism that would perform the potential-reversals at the proper

moments, but the felicitous law of the equality of the intervals makes all easy—all that is needed is to connect an oscillator of the proper frequency (determined by the strength of the magnetic field and the charge and mass of the ions) across the pair of half-boxes.[10]

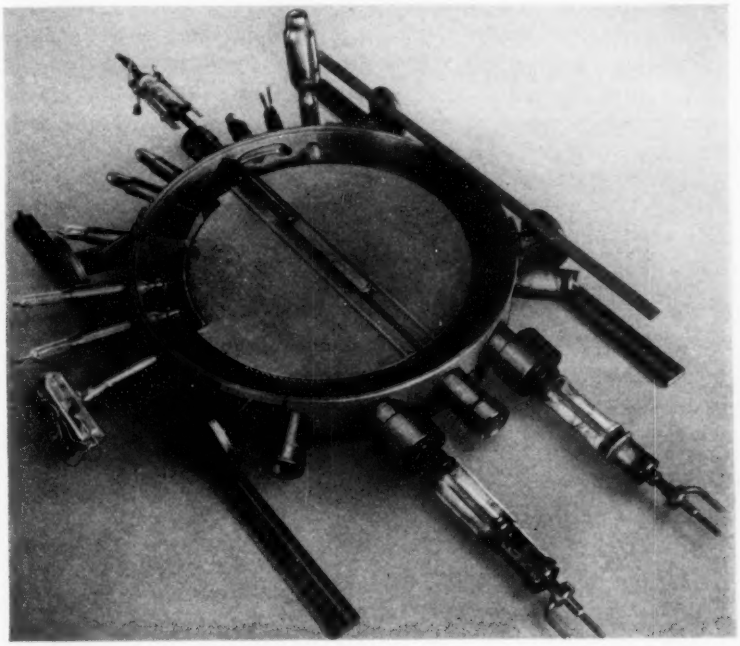


Fig. 4—Photograph of the apparatus sketched in Fig. 3. (E. O. Lawrence)

The sketch of Fig. 3 is of the apparatus wherewith Henderson observed the transmutation of lithium by protons (pp. 140–142). It is filled with hydrogen of a low density, so that electrons proceeding from the hot filament *F* at the center ionize the gas and produce a sufficient number of protons. These are whirled around and around in ever-widening semicircles, till after a number of circuits which may be as high as one hundred and fifty they arrive at the boundary of the

[10] One may do without the magnetic field, arranging to have the ions proceed along a straight line and to accelerate them at definite points along that line, by voltages produced in rhythm by an oscillator; the points of application of the voltages must be spaced according to a particular way, and the apparatus is inconveniently long, being longer the lighter the ion; it has been successfully employed with mercury ions by Lawrence and some of his colleagues.

half-box B opposite the charged electrode D, which deflects them enough to bring them into the cup-shaped receptacle at the far end of which the crystals of lithium fluoride are spread. The fragments of lithium nuclei which are observed are those which escape to the left in such directions as to enter the Geiger counter G. In this apparatus the radius of the outermost circle was 11.5 cm., the magnetic field 14,000 gauss; the potential-difference between the half-boxes never attained as much as 5000 volts, but it was reversed 4.2 millions of

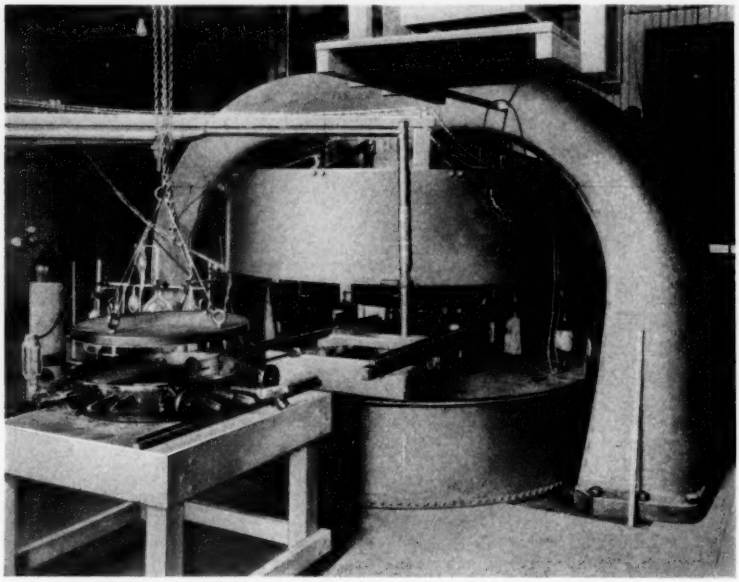


Fig. 5—The Lawrence apparatus for cumulative acceleration, beside the colossal magnet between the poles of which it is placed. (Lawrence)

times in a second, so that after three hundred reversals the protons had arrived at the limit of the box and were ready to strike the lithium nuclei with an energy of 1.23 MEV. By using a bigger pair of half-boxes in a more extensive magnetic field, this energy could be augmented; and by viewing the size of the magnet in Fig. 5 one sees what an augmentation is now imminent. The currents were inferior to those achieved by the apparatus of the Cavendish school, being mostly of the order of a few millimicroamperes (one millimicroampere or $m\mu a = 10^{-3}\ \mu a$).

Detection and Measurement of Transmutation

While thus the scope of transmutation has been so vastly extended in the past eighteen months, there is one limit which has not yet been passed. *No product of transmutation has yet been detected by any chemical means.* Many a plate of metal has been bombarded with protons or with alpha-particles, but no man has seen it change into a plate of another metal, nor alter in any of its chemical properties; many a tubeful of gas has been bombarded, but no man has observed the qualities or the spectrum-lines of another gas appearing in the content of the tube. All that is ever observed is an outpouring of material particles from the piece of bombarded matter; particles of such a nature, that they must come from the nuclei of the atoms. One expects this statement to go out of date from one morning to the next; but at the moment of this writing it is still as true as it was in 1919 when Rutherford first disintegrated nuclei, and broader in one respect only. From 1919 until 1932, one would have said "charged particles"; but since the winter of 1932, it is known that either charged particles or uncharged may be driven out of nuclei, by the appropriate impacts.

Thus there are two great experimental problems, and not one only; beside the problem of producing the streams of bombarding corpuscles, there is that of detecting and of recognizing the particles which fly forth from the bombarded nuclei—the "fragments," I will say. There is a grave objection to this term, and to the common name "disintegration" for the process. Both suggest a picture of the nucleus as a structure of pre-existing pieces which the impact breaks apart and scatters. This picture is surely incorrect, for there are cases in which the fragments contain the susbtance of the impinging corpuscles. In fact, if we define "fragment"—as we should—to include the part which in most of the experiments does not escape from the target bulk, we may say that this kind of case is frequent, and perhaps indeed that there is no other kind! Nevertheless we seem to be unable to get along without the words "disintegration" and "fragment."

For detecting protons and more massive fragments which are charged, there are three methods.

The *first method* (A) is that of observing the scintillations, which fast charged particles produce when they impinge on fluorescent screens. This is the classic and historic method, by which were made the earliest proofs of transmutation by impact of alpha-particles (which I described at length in the earlier article) and also the earliest proof of transmutation by protons. Of late years this method has been largely displaced by the others. Few people outside of the Cavendish

Laboratory and the Institut für Radiumforschung in Vienna have ever submitted themselves to the long, tedious and nerve-racking process of counting thousands of dim flashes for periods of hours in darkened rooms with dark-adapted eyes; and if two disagreed as to what was observed, there was no objective way of deciding between them. The newer methods abolish this strain; they can readily be so shaped as to leave a permanent record, which anyone may consult and analyze for himself; and they are capable of measuring the ionizing power of the fragments. Nevertheless the eldest method still retains the unique advantage that no barrier whatever, not even a gas, need intervene between the detecting screen and the source of the fragments; and also it is often employed by those accustomed to scintillations as a check upon the others.

The *second method* (B) is that of the expansion-chamber or cloud-chamber of C. T. R. Wilson, whereby the tracks of ionizing particles across a gas are made visible by droplets of water which condense upon the ions. This is the splendid invention which is the joy of all who write or lecture on atomic physics, since it enables them to decorate their exposition with pictures which make real the things of which they speak. It has virtue for the investigator also, especially since it may show in a single vivid photograph how many fragments there are formed in a single process, what are the directions in which they fly away, and how far they are able to travel through the gas. The curvature of the track in an applied magnetic field supplies the value of the momentum of the particle which made the track, if the nature of the particle be known; and this last may often be guessed from the aspect of the track, or assured by independent data. The major disadvantage of the method is, that the apparatus records only the particles which fly off during about a hundredth of a second, and then lies idle for several seconds or even minutes while it is being prepared for its next brief interval of effectiveness.

The *third method* (C)—or group of methods rather, for the variants are legion—is the detection by purely electrical methods of the ions which the fragments produce as they shoot across the gas of an ionization-chamber. A fast-flying charged particle loses on the average 30 to 35 electron-volts for every ion, or rather every ion-pair, which it produces.[11] To see the utility of this theorem, turn it around; the number of ion-pairs produced by a fast charged particle going through a gas is about a thirtieth of the number of electron volts which it loses in its transit. A fast alpha-particle, such as are spontaneously emitted by radon, or constitute the fragments springing out

[11] "Electrical Phenomena in Gases," pp. 52, 70–71.

of lithium bombarded by protons, has about eight million electron-volts; if it enters an ionization-chamber filled with gas so dense that it is brought completely to a stop, the ion-pairs appearing are about a quarter of a million. The upper limit occurring in practice is possibly twice as high, but is very rarely met with; there is no lower limit, but every incentive to push downward and ever downward the least amount of ionization which can be detected.

Twenty-five years ago, it would have been impossible to detect by electrical means so few as a quarter of a million ions. (The total number produced *e.g.* by an alpha-particle was determined by measuring the total ionization produced by a known and very great number of particles.) This problem was however destined to be solved in many ways, which I will group under four headings:

(C1) By arranging to have each particle touch off a brief but violent discharge, something like an invisible spark, in the gas of the ionization-chamber. There is a strong electric field applied between the electrodes of the chamber, whereby the "primary" ions which the particle forms as it travels across the gas are caused to produce (directly and indirectly) vast numbers of extra or "secondary" ions; and these suffice to make a sensible effect in the external circuit. The idea was first put into practice by Rutherford and Geiger in 1908, and the scheme is commonly known by Geiger's name. One of the electrodes must be either a fairly sharp point or a fairly thin wire, and there are a number of empirical rules (some partially understood, some not at all) about the size and shape of the chamber, the proportioning and the conditioning of the electrodes, the nature and the purity and the density of the gas, and the magnitude of the field. The voltage across the gas must lie within a definite range, often pretty narrow; if it is lower the particles do not produce discharges, if it is higher a single discharge may last indefinitely. The ratio of the number of secondary to the number of primary ions is usually not constant and usually not measured; most of the various forms of the device serve solely to detect or count the particles, and they are known as "Geiger counters." Often a loudspeaker is connected into the circuit of the ionization-chamber, and each discharge produces an audible clack, so that by the Geiger method one hears the passage of a corpuscle as by the Wilson method one sees it. Sometimes the discharges are recorded and the record examined at leisure.

(C2) By modifying the foregoing scheme so that the number of secondary ions shall be proportional to the number of primary ions, and a measurement of their total charge shall give at least a relative value of the ionizing-power of the traversing particle. This is a

recent achievement of Geiger and Klemperer. The process may be called *internal amplification* of the primary ionization, the amplification being in a constant proportion, or, as people carelessly call it, "linear."

(C3) By developing an electrometer or electroscope so sensitive that it is able to detect and even measure the total charge of a few thousands of ions, without amplification. This was first achieved, or at any rate applied to transmutation, by G. Hoffmann of Halle, and his associate Pose; the latter was able to observe fragments of aluminium nuclei (ejected by alpha-particles) which produced as few as three thousand ion-pairs. The major difficulty seems to be, that the electroscope takes a large fraction of a minute to perform its deflection and then recover its readiness to respond to another particle. Pose in his experiments observed only some thirty fragments to the hour.[12]

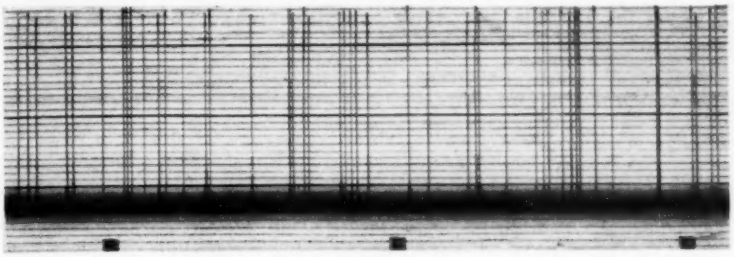
(C4) By applying *external amplification* to the feeble impulse which the primary ions due to a single particle produce in the external circuit, and which is imperceptible to an electroscope of normal and convenient quickness of response. This is done by developing the superb techniques of amplification which modern vacuum-tubes have rendered feasible, and like the three foregoing schemes is an achievement of the last few years, having been carried on especially by Wynn-Williams of the Cavendish Laboratory and Dunning of Columbia.

I show as Fig. 6 three records made with Dunning's apparatus, wherein every vertical line is due to an ionizing particle, and is proportional in length to the number of ions which the particle produced in crossing a shallow chamber.[13] The lines of great and nearly uniform length which appear in record (*a*) are due to alpha-particles from polonium; these all had nearly the same speed and were moving in nearly parallel lines when they entered the chamber, and it is evident that in crossing the gas they all made nearly the same amount of ionization; they left with a good deal of their initial kinetic energy unspent. The lines in record (*b*) are caused by protons; their diversity in length is chiefly due to the wide variety of speeds which the protons had when they entered the chamber, for these were fragments of the disintegration of aluminium by alpha-particles, and therefore had a broad distribution-in-speed (page 147). As these words imply, and as I will stress presently, the ionization produced by a charged particle

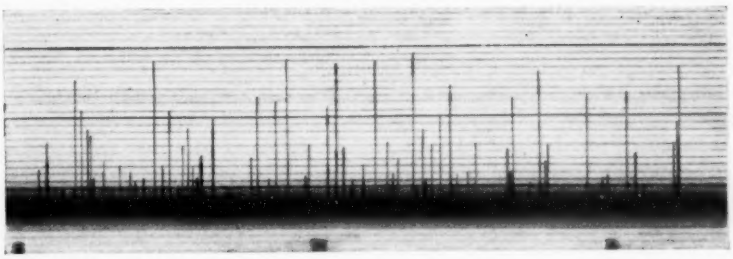
[12] It deserves to be recorded that in their blank experiments, Hoffmann and Pose during one research observed deflections at the average rate of 1.22 per hour, but observed altogether 197 of them!

[13] I am much indebted to Dr. Dunning for these pictures, made especially for this article. He writes of (*b*): "The minimum amount of ionization detectable here is well under 1000 ions; probably it could be pushed down to 250 ions." Consecutive dots at the bottom of each record mark off the minutes.

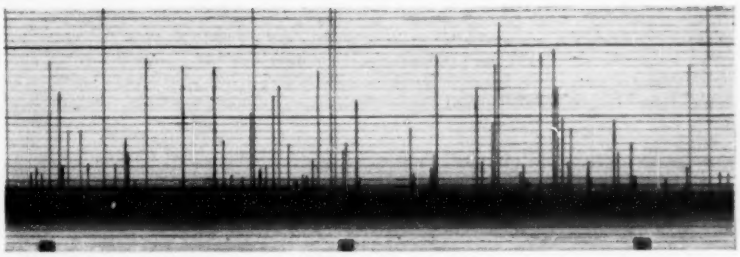
of given kind depends on its speed; the greatest amount which (in this particular chamber) a proton could ever produce, with its most favorable speed, is indicated by the longest lines in (*b*), and one sees that even these are definitely shorter than the lines in (*a*) due to



(*a*)



(*b*)



(*c*)

Fig. 6—Three records of the ionization produced by individual particles in a shallow ionization-chamber: (*a*) alpha-particles of nearly the same speed, (*b*) protons of various speeds, (*c*) particles of several kinds which had been set into motion by impacts of neutrons. The fogging along the base-lines is much fainter in the original records than in these reproductions. (J. R. Dunning)

alpha-particles.[14] The lines in record (*c*) were obtained when neutrons were traversing the chamber and a piece of paraffin outside of it; not they, but the charged nuclei which they strike and impel, are producing the record. Those lines which are longer than any in (*b*) are certainly not due to protons; they must be caused by recoiling nuclei of the atoms of the gas which fills the chamber (air), and which have various speeds because the neutrons strike them more or less glancingly (and probably do not themselves all have the same speed). The shorter lines are due in part to such nuclei, chiefly to protons ejected from the paraffin in such directions that they cross the chamber. Some are very short indeed, half-lost in the dusky haze due to the perpetual wiggling of the oscillograph mirror caused by gamma-rays; they are made by the fastest of the protons. Observations by expansion-chambers and with applied magnetic fields have proved this classification of the particles.

All of these methods are available for detecting charged particles which are protons or alpha-particles or corpuscles of a yet greater mass than these. For electrons the problem is harder.

An electron of given energy—say x thousands of electron-volts—is able to make roughly as many ion-pairs in a gas as could a proton or an alpha-particle of equal energy: that is to say, about $30x$. Nevertheless it produces much less ionization in an ordinary chamber than either of these last. This seeming paradox is due to the facts that the ion-pairs produced by the electron are relatively far apart and the loss of energy per centimeter of path is correspondingly low, so that in an ionization-chamber of reasonable dimensions and customary density of gas the traversing electron produces only a few hundreds or perhaps one or two thousands of ion-pairs before it reaches the opposite side of the chamber and plunges into the wall.

This is made evident by the Wilson method, the tracks of electrons appearing much thinner—less richly peopled with droplets, that is to say—than those of alpha-particles or protons. The expansion-chamber therefore is available for observing fast electrons, and so to a certain extent is the Geiger counter, which skilful observers can adjust so that it will react to these bodies. None of the other methods has yet been used with success. The ions produced by a single electron in an ionization-chamber are apparently too few to observe without amplification or even to amplify successfully, and the scintillations too faint. If one has neither expansion-chamber nor Geiger counter available, the only thing to be done is to measure the total ionization

[14] The contrast is much more striking than the records suggest, for the amplification was fourfold greater when (*b*) and (*c*) were made than when (*a*) was made.

produced by great numbers of electrons, and attempt to estimate these numbers. This is done in the study of the beta-rays or fast electrons emitted from radioactive nuclei, and in the study of cosmic rays; but the method has not yet been applied to the rays emitted from atoms undergoing transmutation by impacts, and apart from Joliot's observations on positive electrons (page 102 *supra*) nothing yet is known of any electrons which may be emitted by these.

Since individual electrons are so difficult or impossible to observe by the customary methods, one might suppose that at any rate they never annoy the observer. This unluckily is not so; for if electrons are numerous, they may keep the electrometer needle (in the method C4, for example) in a perpetual tremor, producing a so-called "background" over which even the strong sharp impulses due to alpha-particles or protons may fail to stand out. It is even possible for a chance coincidence or near-coincidence of several electrons to make a record which cannot be disinguished from that of a single particle of greater ionizing power. The scintillation-method suffers from a like defect, for if the fluorescent screen is heavily bombarded with electrons —or with gamma-rays, which liberate electrons from the fluorescent stuff and the surrounding matter—it shines all over with a feeble glow, against which the flashes made by more massive ions are difficult to discern. The most casual student of transmutation cannot fail to notice that polonium is generally used, of recent years, as the source of alpha-particles for bombardment. Probably he infers that either it is especially abundant or else supplies especially fast particles. But in both respects polonium is inferior to another customary source, radon mixed with its descendants radium A and radium B. It is used because it emits no gamma-rays but feeble ones of low penetration, whereas the other source pours out abundant and powerful photons which flood any nearby ionization-chamber with electrons and confuse the electrometer. Dunning's amplifying circuit, whereby he detected charged nuclei set into motion by neutrons, was so devised as to discriminate against the feeble but many impulses produced by these electrons and in favor of the occasional stronger ones produced by the massive particles; and this device enabled him to use a source of the latter type providing fifty times as many alpha-particles to engender the neutrons, as the largest amount of polonium ever employed.

Neutrons, I recall, are detected by observing the protons and more massive nuclei which they convert by impact into fast-flying ionizing particles, and photons by observing the electrons on which they have the like effect; the problems of getting the data are thus not new, it is the problem of interpreting them which is changed.

The next important question is, how the fragments are identified as protons, or as alpha-particles, or otherwise, from the data. Few as yet are the cases in which the identification is full and undeniable. In the earlier paper I described Stetter's measurements of charge-to-mass ratio for the fragments produced by impacts of alpha-particles against boron, carbon, fluorine and aluminium, which gave values identical with that for protons within the observational uncertainty of five per cent. As for the fragments produced by impacts of protons, the best direct evidence is that which appears in Fig. 7. Cockcroft

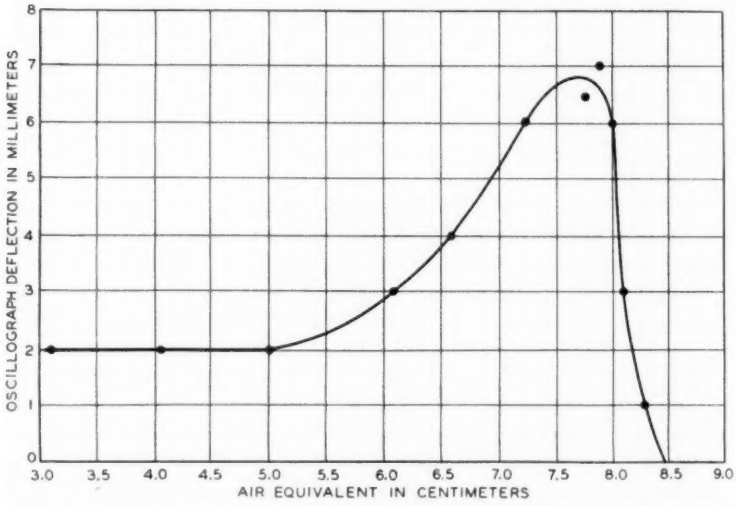


Fig. 7—Ionization produced in a shallow chamber by fragments (of the transmutation of lithium by protons) which have passed through screens of various thicknesses. (Cockcroft and Walton)

and Walton had an ionization-chamber only 3 mm. across, and the fragments from bombarded lithium traversed it completely, producing a few thousand ion-pairs apiece which were detected and measured with the aid of an amplifying circuit of Wynn-Williams according to the method C4. When mica sheets were interposed in the path of the fragments from the lithium, they were slowed down but still kept energy enough (so long as the sheets were not too thick altogether) to travel across the chamber; and the curve of Fig. 7 represents the number of ion-pairs produced per fragment, as function of a quantity x proportional to the thickness of mica which the fragments have

traversed ("air-equivalent" of the mica, p. 127 *infra*).[15] The point is, that exactly the same curve was obtained when a beam of alpha-particles was projected through the same thicknesses of mica into the same chamber. Mere similarity in the shape of the curves would prove nothing, for this is the shape obtained with all kinds of charged particles, electrons and protons and more massive charged nuclei; in particular, every such curve rises from zero to a maximum and thereafter descends continually as the energy of the particles is raised indefinitely upward from the least value sufficient for ionization.[16] However, the ordinates of the curve of Fig. 7 are *equal* to those of the alpha-particle curve, and about four times as great as would have been observed with protons; and this it is which proves the fragments to be alpha-particles. Almost as good a proof could be made by two measurements: by measuring the range of the fragments and the total ionization produced by any fragment in a chamber deep enough to swallow it up, and comparing the latter datum with the ionization produced in the same chamber by an alpha-particle of equal range. This proof, or some other substantially like it, has been adduced in certain cases. When alpha-particles are the agents of the transmutation, the same test has proved in several cases that the fragments are protons. In some cases the test has not yet been applied.

I have already had to speak of interposing mica in the path of the fragments, in order to learn something about them. This is a procedure with which it is necessary to be familiar. It would be very pleasant indeed to be able to apply electric and magnetic deflecting fields to a narrow stream of fragments all flying in the same direction, for one could then spread it out into a velocity-spectrum, and not only identify the corpuscles perfectly but also determine their distribution-in-range, which as we shall presently see is of the first importance. This has not yet been done, partly (I presume) because of the high fieldstrengths that would be needed, chiefly because the available streams of particles are too scanty. It will be a happy day when at last we get streams of fragments so intense that they can be dispersed into a velocity-spectrum which will appear imprinted on a photographic film, as has been feasible for years with beta-rays. For the time being we must be content with curves such as many figures in this article display, Figs. 8 and 9 and 11 for example.

[15] The quantity plotted as ordinate is obtained from such records as those of Fig. 6, in which every fragment produces a vertical line. Cockcroft and Walton observed many such lines for each thickness of mica, and ascertained in each case the most frequently-occurring value of line-length.

[16] "Electrical Phenomena in Gases," pp. 40–44, 70–71. Such a curve as that of Fig. 7 is sometimes called a "Bragg curve."

These are curves in which the abscissa stands for the thickness of a special kind of matter (air of a standard density) interposed in the path of the fragments, and the ordinate for the number of fragments detected on the far side of that matter; I will call them "integral distribution-in-range" curves representing the number $f(x)$ of particles able to traverse thickness x. Were one to differentiate them, one would get the "differential distribution-in-range" curves, representing a function $f'(x)$ such that $f'(x)dx$ stands for the number of particles able to traverse thickness x but not additional thickness dx—the particles which are said to have "ranges" between x and $x + dx$.

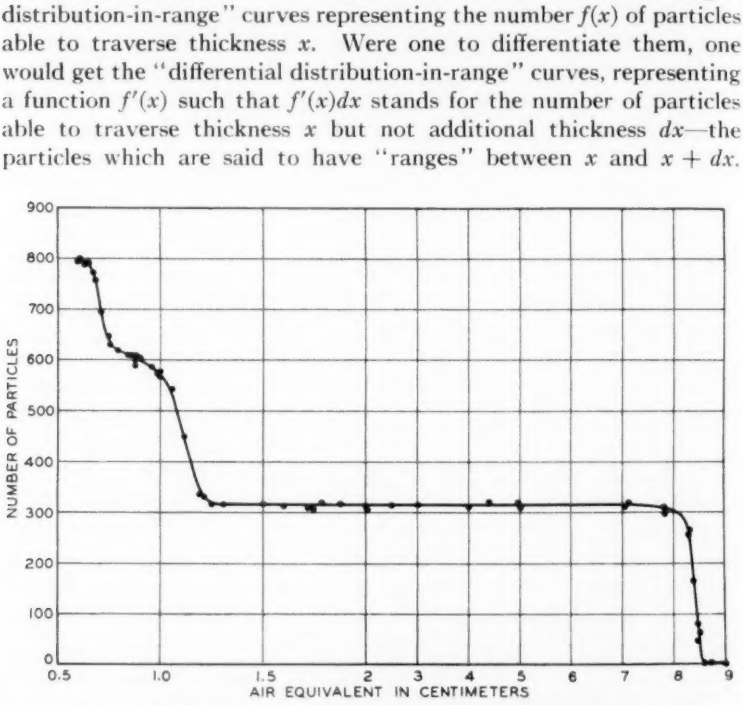


Fig. 8—Integral distribution-in-range curve of the fragments resulting from bombardment of lithium by protons. (Oliphant Kinsey & Rutherford)

These, however, are usually not plotted,[17] and one must accustom himself to draw the proper inferences from the integral curves.

The clearest of these to read are those which are shaped like a staircase, with steep rises connecting horizontal parts called paliers or plateaux. A steep rise extending over a narrow interval of x signifies a "group" of fragments all having ranges close together. A plateau extending over a broad interval of x signifies that no particle has a range comprised anywhere in this interval. An integral curve in the form of a staircase therefore implies the analogue of a line-spectrum,

[17] One of the rare examples is reproduced in "Transmutation," *B. S. T. J.*, Vol. X, p. 650 (Oct. 1931), from the work of Bothe and Fränz.

the particles being classifiable into groups each with its characteristic speed. But if in such a curve there is a long sloping arc (as in Fig. 9), it implies the analogue of a continuous spectrum, there being particles of all ranges over a notable interval.

The "stopping" or "absorbing" screens which are used in determining these curves are usually sheets of mica or of aluminium. The curves are not however plotted against the actual thickness of the interposed strata of mica or whatever else the substance may be, but against the "air-equivalent" or thickness of the stratum of air of standard density [18] which is known by separate experiments to have the same effect in slowing down and stopping charged particles, the

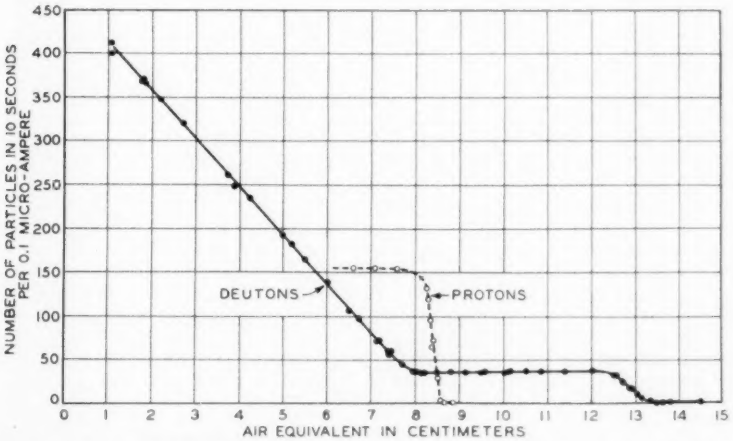


Fig. 9—Integral distribution-in-range curve of the fragments resulting from bombardment of lithium by deutons. (Oliphant Kinsey & Rutherford)

same "stopping-power." It is the air-equivalent which is the quantity x of the preceding paragraphs and the abscissa (often termed "absorption") of Fig. 8 and nearly all other such figures. The ratio between the actual thickness of a layer of matter and the equivalent thickness of air is roughly (but only roughly) the reciprocal of the ratio of their densities. The sheets of metal or of mica used in the experiments are therefore very thin (it has been possible to make screens of mica so tenuous that their air-equivalent is only 0.15 mm.) and the thinnest must be bolstered up by stiff metal grids, of which the wires block a considerable fraction of the beam. It is also possible

[18] There are unluckily two standards of density, one being that of air at 0° C. and 760 mm. Hg, the other that of air at 15° C. and 760 mm. Hg; see "Transmutation," footnote on p. 643, *B. S. T. J.*, Oct. 1931. The latter is used in this article.

to use air (or some other gas) of adjustable density; when the scintillation-method is employed, the gas may fill the entire space between the source of the fragments and the fluorescent screen; with other methods of detecting the fragments, it must be contained in a cell which the stream enters and leaves through windows of mica or similar substance.

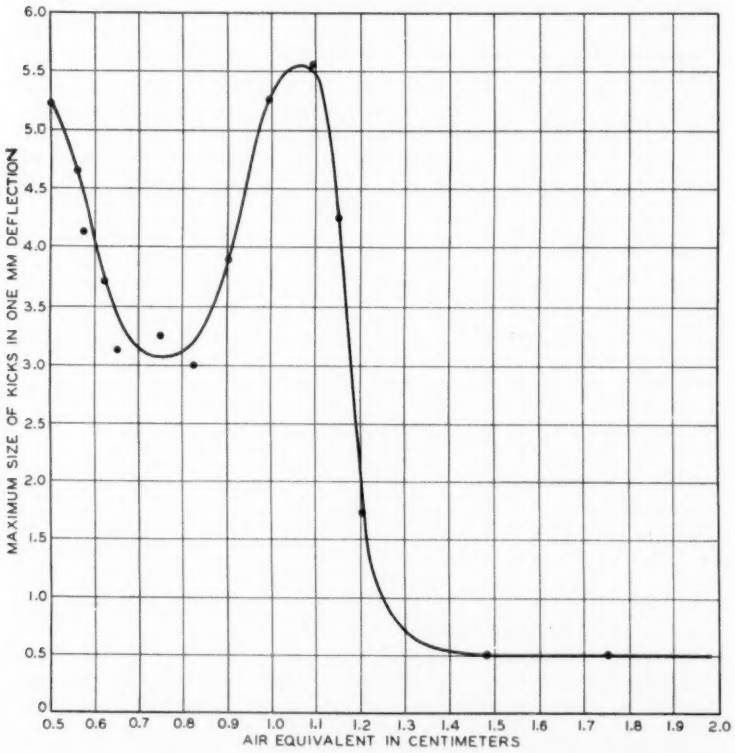


Fig. 10—Ionization produced in a shallow chamber by the least penetrating fragments from the transmutation of lithium by protons. (Oliphant Kinsey & Rutherford)

There is an interesting and important way of confirming the steps in an integral curve such as those of Figs. 8 and 9. Near the rise of such a step, the thickness of the intercepting matter is such that many particles are approaching the ends of their ranges when they emerge from the last of the screens. Suppose that this last screen is adjoined by a very thin ionization-chamber, like that with which the curve

of Fig. 9 was obtained. Let the air-equivalent x of the total thickness of the screens be varied, and let the average number of ions produced per particle in the chamber be measured and plotted as function of x. Recalling Fig. 7 and what was said in respect to it, the reader will see that the resulting curve should have a peak wherever the integral distribution-in-range curve has a step. This has been verified several times, and there are cases in which these peaks have been taken as

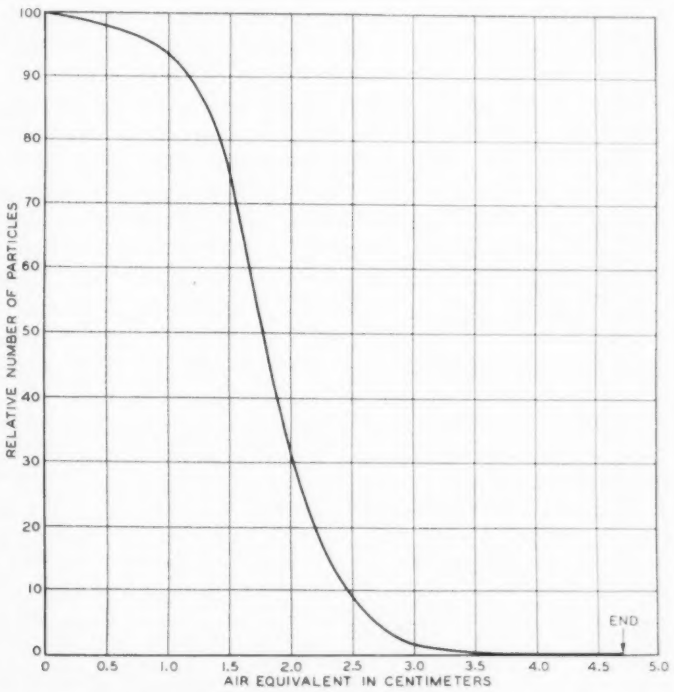


Fig. 11—Integral distribution-in-range curve of the fragments resulting from the bombardment of boron by protons. (Oliphant & Rutherford)

clearer evidence for the existence of groups than the shape of the integral curve itself (Fig. 10). Peaks may also appear in a curve of which the ordinate is the *total* ionization produced in the very thin chamber by all the fragments which enter it.

Anyone at all acquainted with physical experiments will readily suspect that the steps of actual integral curves "ought to be" steeper than they are. I mean: that he will form the hypothesis that perpen-

dicular rises would be observed instead of rounded-off and sloping ones, if only the pencil of fragments passing through the absorbers were ideally narrow and cylindrical, and were produced by bombardment of atoms with particles all of the same speed; and he will attribute the rounding-off of the steps to the facts that the fragments actually form a divergent and conical beam, and the atoms from which they come have been struck by impinging particles of diverse speeds. This idea is strongly supported by the facts that the steps are notably steepened when the divergence or "aperture" of the beam of fragments is reduced, and when the diversity of speeds among the bombarding particles is narrowed.

The former of these variables is controlled by the slits and diaphragms which bound the beam, and the latter by the thickness of the bombarded target whence the fragments proceed; for the bombarding particles are slowed down as they dive deeper into the target, and nuclei at different depths receive impacts of different energy, and thus there is a wider diversity of speeds among the particles when they finally make their impacts than there is among them when they start from their source. But as one cuts down either the thickness of the target or the aperture of the beam of fragments, one reduces the number of fragments which come to the detecting apparatus, and reaches a limit when this number becomes too small to be observed in any convenient time. Progress in approaching ideal conditions therefore depends on progress in multiplying the number of fragments by multiplying the strength of the bombarding beam. We may count on a yet greater steepening of such steps as those of Fig. 8, when the enormous streams of bombarding protons produced by Oliphant and Rutherford are applied to very thin films and the distribution-in-range of the resulting fragments is measured. A corresponding improvement of the curves obtained when alpha-particles are the bombarders is still in the not-immediate future. Whether under ideal conditions the steps would be absolutely perpendicular, and all the fragments of a group have exactly the same speeds, is not as yet to be safely inferred from the data.

There remains the great problem of converting distribution-in-range curves into distribution-in-speed or distribution-in-energy curves, and thus determining the energy or the speed of fragments belonging to a group of which the range is known. The recent developments of research in transmutation and in cosmic rays have elevated this to the rank of the major problems of physics. For alpha-particles of ranges of 8.6 cm. and less, it is practically solved by empirical means; for such alpha-particles are supplied in such abundance by radioactive

bodies that it has already been feasible to measure by deflection-methods the speeds corresponding to a large number of different ranges, and plot an empirical speed-*vs*-range curve which is fixed by so many points of observation that there is no important uncertainty in making interpolation between these. For alpha-particles of range superior to 8.6 cm., such as often occur among fragments of transmutation, it has heretofore been necessary to extrapolate; but very lately the empirical curve has been extended onward to 11.6 cm., thanks to a powerful new magnet at the Cavendish which is able to deflect the paths of alpha-particles of even such rapidity.[19] With protons our knowledge of the range-*vs*-energy relation is less extensive and less accurate, and an improvement thereof should be one of the first and most important by-products of the new methods for imparting high energies to ions. For charged nuclei of other elements than hydrogen and helium, relatively little is assured (what is known has been found out chiefly by Blackett and his school)[20]; but this lack has not as yet been much of an impediment to the study of transmutation, except in certain cases involving impacts by neutrons.

Transmutation by Impacts of Protons and Deutons

The earliest element to be transmuted by protons in the laboratory —indeed the first to be transmuted by man with any agent other than the alpha-particle—was lithium. It was fortunate that Cockcroft and Walton began with this element, for its behavior turned out to be uniquely lucid. In most disintegrations, a single fragment is detected, and there must be a massive residue which remains unseen, staying hid within the substance of the bombarded target. But in some at least of the transformations which occur when lithium nuclei are struck by protons or deutons, there seems to be no hidden residue; every fragment is observed and recognized. These are processes of "nuclear chemistry" of which we fully discern both the beginning and the end; and they are described by the quasi-chemical equations:

[19] Rutherford et al., *Proc. Roy. Soc.* **139**, 617–637 (1933). The empirical curve departs slightly from a third-power law (range proportional to cube of speed) and the results are expressed by an empirical formula for the departure. See also G. H. Briggs, *Proc. Roy. Soc.* **139**, 638–659 (1933).

[20] See N. Feather, *Proc. Roy. Soc.* **141**, 204 (1933) and literature there cited. The observations are made upon tracks which appear in Wilson chambers when the contained gas is bombarded by alpha-particles, and which are the tracks of objects of atomic mass that have suffered violent impacts. It is presumed (though not always proved) that these objects are solitary or "bare" nuclei, not accompanied by any of the orbital electrons which attended them before the impacts. Some (but not all) of the data conform to the empirical rule that the ratio of the ranges of two nuclei of masses m_1 and m_2 and of charges Z_1e and Z_2e, when the two have the same speed, is $(m_1/m_2)(Z_1/Z_2)^{1/2}$.

$$_1H^1 + {}_3Li^7 + T_0 = 2{}_2He^4 + T_1, \tag{1}$$

$$_1H^2 + {}_3Li^6 + T_0 = 2{}_2He^4 + T_1, \tag{2}$$

of which the first has already appeared in Part I. of this article.

These are to be regarded as equations for mass and energy, owing to the equivalence of these two entities. Attached to the symbol of each atom are its mass-number as superscript and its atomic number as subscript (and, incidentally, every such equation must balance when considered as an ordinary equation in either the mass-numbers or the atomic numbers). The symbols T_0 and T_1 stand for the total kinetic energy of the particles *before* and the particles *after* the transmutation, expressed in mass-units. (I recall from Part I. that a mass-unit is one-sixteenth the mass of an ${}_8O^{16}$ atom, and that one million electron-volts is equal to 0.00107 of one mass-unit.) The other symbols then stand for the rest-masses of the nuclei of the atoms in question. It would be proper, and in accordance with the spirit of relativity, to leave out the symbols T_0 and T_1 and consider each of the other symbols as standing for the *total* mass of the nucleus, viz. the sum of its rest-mass and the extra mass resulting from its speed. When hereinafter the symbols T_0 and T_1 are absent from such an equation, the others are thus to be interpreted.

The suggestion thus is, that when a proton meets with a ${}_3Li^7$ nucleus or a deuton with a ${}_3Li^6$ nucleus, either process ends in the formation of two helium nuclei—alpha-particles—out of the substance of the original bodies. It is further suggested that these nuclei share kinetic energy amounting to T_1; and if they are emitted in directions making equal angles with that of the impinging particles—the "symmetrical case" which (as we shall see) is most commonly observed—they must share T_1 equally in order to assure conservation of momentum. Now the rest-masses of all the nuclei figuring in equations (1) and (2) are accurately known through the work of Aston and of Bainbridge. Taking them from Table I and substituting them into the equations, and using the electron-volt for our unit, we get:

$$T_1 = T_0 + 16.8 \cdot 10^6, \tag{3}$$

$$T_1 = T_0 + 22.2 \cdot 10^6, \tag{4}$$

in the two cases,[21] and therefore expect alpha-particles paired with one another, their kinetic energies amounting altogether to these values.

[21] For these numerical values and their uncertainties, see K. T. Bainbridge, *Phys. Rev.* (2), **44**, 123 (July 15, 1933).

It is the verification of these predictions which gives us such great confidence that we have recognized the processes which really happen.

I have already said how Cockcroft and Walton proved that the fragments, when lithium is bombarded by protons, are alpha-particles. The integral distribution-in-range curve of these fragments, obtained by Oliphant Kinsey and Rutherford with the apparatus of Fig. 2 and proton-currents running up to $50\mu a$, appears in Fig. 8; and that for the fragments created when deutons are used instead of protons appears in Fig. 9. In both of these one cannot but be struck by the beautiful long horizontal plateaux, and the sharpness of the steps which end them on the right. The groups of fragments of which these steps are the signs have ranges stated by the observers as 8.4 and 13.2 cm respectively, with uncertainties of $\pm$ 0.2 cm. (These figures are evidently taken from the bottom of the step, probably because it is assumed that under ideal conditions of narrow beam and thin bombarded film—the actual beam had a divergence of about 15° and the actual target was thick—the step would rise vertically from the point whence it actually begins to rise obliquely.) The corresponding energy-values are estimated as 8.6 and 11.5 MEV (millions of electron-volts) respectively; and as T_0, the energy of the impinging protons, is at most two-tenths of a million, these values may be compared directly with the halves of the numbers in equations (3) and (4). Meanwhile at Berkeley, Lewis Livingston and Lawrence were driving deutons with an energy of 1.33 MEV—no longer negligible—against lithium, and observing fragments with a range of 14.8 cm., corresponding to an energy of 12.5 MEV; and this is to be compared with half of 23.7 millions on the right-hand side of equation (4).

The agreement in the case of protons impinging on lithium is admirable, and well within the uncertainty of the data. The agreements in the cases of deutons impinging on lithium are ostensibly not so good, but this is not so serious as it seems at first glance, because of the required extrapolation of the range-*vs*-energy curve of alpha-particles (page 131), and because it is not always the "symmetrical case" which occurs. For the present there is no compelling reason to suppose that equation (2) is contradicted by the data.

A further point susceptible of test: if the processes described by equations (1) and (2) are actual, the alpha-particles of the stated ranges must be shot off in pairs, the two members of each pair flying off in almost opposite directions—in directions which would be exactly opposite were it not for the original momentum of the proton, but which because of that momentum must make with one another an angle slightly (and calculably) less than 180°. Cockcroft and Walton

made the test with a pair of Geiger counters set on opposite sides of the bombarded lithium, and got a positive result; but it is the expansion-chamber which is suited by its nature for supplying the most magnificent of proofs. To achieve this, one must put the bombarded target of lithium in the middle of the chamber, and photograph the tracks from above; and since the bombarding stream must come through vacuum while the chamber must be filled with moistened air, the target must be separated from the air by walls of mica thick

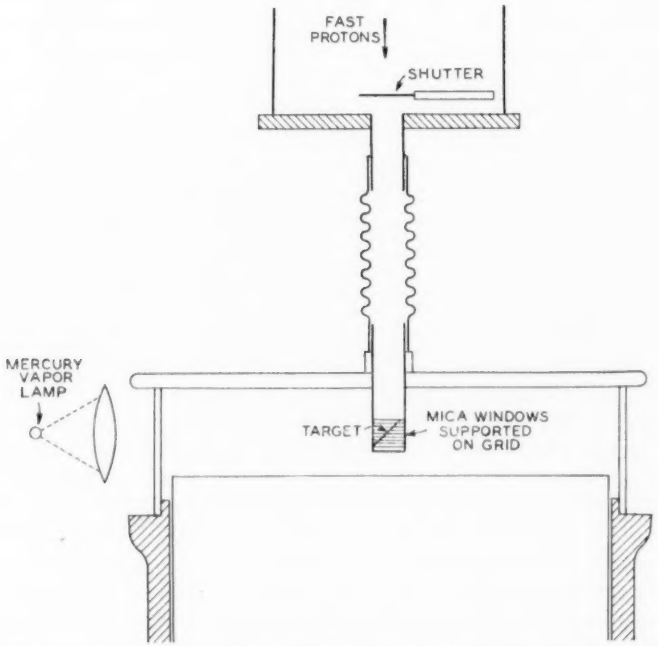


Fig. 12—Diagram of arrangement for observing tracks of fragments by the expansion-method. (After Dee and Walton)

enough to withstand the pressure and thin enough to let the fragments pass. The scheme is clearly depicted in Fig. 12. One notices that the design is such that the pairs which are observed are those of which the directions are nearly at right angles to the proton-beam—the "symmetrical case" aforesaid.

This experiment was first performed by Kirchner of Munich, who got several pictures of paired fragments from lithium bombarded by protons. Fig. 13 shows an example. (The third track is rather

annoying, but it was quite an achievement so to adjust the conditions as to get so few as three.) Many splendid examples have lately been published by Dee and Walton of the Cavendish, and Fig. 14 is outstanding among them because the bombarding stream was a mixture of protons and deutons, and the picture shows two pairs of fragments, one apparently due to each of the processes which I have been describing. Those of the pair marked b_1b_2 have the range of 8.4 cm. agreeing with equation (1), while those marked a_1a_2 go definitely farther and even escape from the chamber, which makes it impossible to measure their ranges. Dee and Walton therefore made the walls of the target-capsule thicker, so that more of the energy of the frag-

Fig. 13—Tracks of paired fragments, He nuclei resulting from impact of a proton on a Li^7 nucleus. (Kirchner; *Bäyrische Akademie*)

ments should be consumed in them; the pairs which were obtained with bombarding deutons now ended in the chamber and in the field of view, and their ranges agreed with the 13.2 cm. obtained from the curve of Fig. 9. At least two more of these pairs appear in Fig. 15. Verification of a theory could scarcely go further or be more vivid! Yet there is the additional point, that Kirchner found the angle between the paired paths in his pictures to differ from 180° by just about the amount required by the momentum of the proton.

However not every fragment observed when lithium is bombarded, either by protons or by deutons, results from these superbly simple interactions. Notice in Fig. 8 the two very much rounded steps, suggesting groups of short ranges (1.15 cm. and 0.65 cm.); these are confirmed by the maxima in the curve of Fig. 10 which has already

been explained (page 129). Only tentative theories of these have been made, and it would be of little use to expound them here.[22] Notice then in Fig. 9 the beautiful long *sloping* line adjoining the plateau, and implying a continuous distribution over a wide interval of ranges extending up to 7.8 cm. The numerous shorter tracks of Fig. 15 are due to particles belonging to this continuum. Observe last the integral distribution-in-range curve for the fragments from

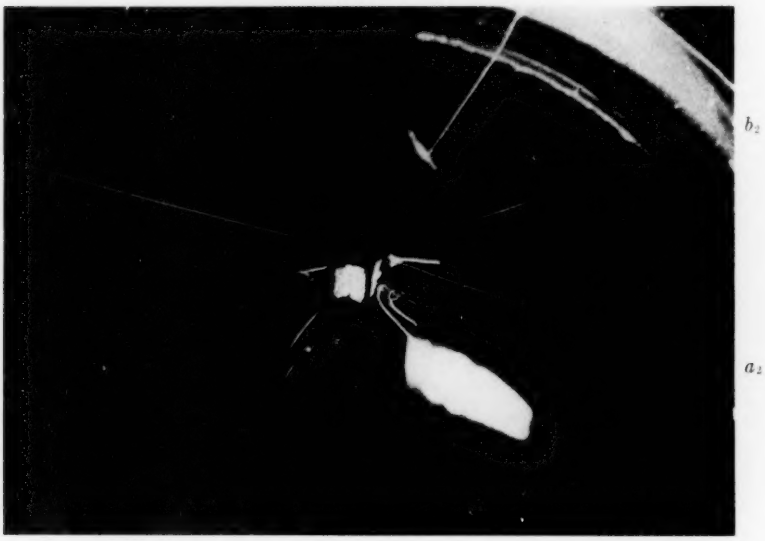


Fig. 14—Tracks of paired fragments, He nuclei believed to result from impact of a proton on a Li^7 nucleus and from impact of a deuton on a Li^6 nucleus. (Dee and Walton; *Proceedings of the Royal Society*)

boron bombarded by protons, Fig. 11; notice that it displays no definite step, but consists of a single sloping arc implying a continuum extending to an upper limit, which on a magnified curve is found to be at 4.7 cm.

It is now suggested that in both of these two last cases we have processes in which there are not two, but three final fragments:

$$ {}_1H^1 + {}_3Li^7 + T_0 = 2{}_2He^4 + {}_0n^1 + T_1, \tag{5}$$

$$ {}_1H^2 + {}_5B^{11} + T_0 = 3{}_2He^4 + T_1, \tag{6}$$

[22] Dee has just announced (*Nature*, **132**, 818–819; Nov. 25, 1933) that these short-range fragments are frequently paired. In doing the experiment he admitted the primary protons into the expansion-chamber through a thin mica window, the target being within.

the symbol ${}_0n^1$ in equation (5) standing for a neutron. When there are three fragments, conservation of momentum no longer demands that the available energy be equally divided among the three, but admits of an infinity of distributions. It is not difficult to find the highest fraction of T_1 which either of the two alpha-particles in case (5), or any of the three in case (6), may receive; this amounts to very nearly one-half in the former, to two-thirds in the latter case.

In equation (5) the rest-masses of all the charged nuclei are known; that of the neutron is still subject to some controversy, but if we

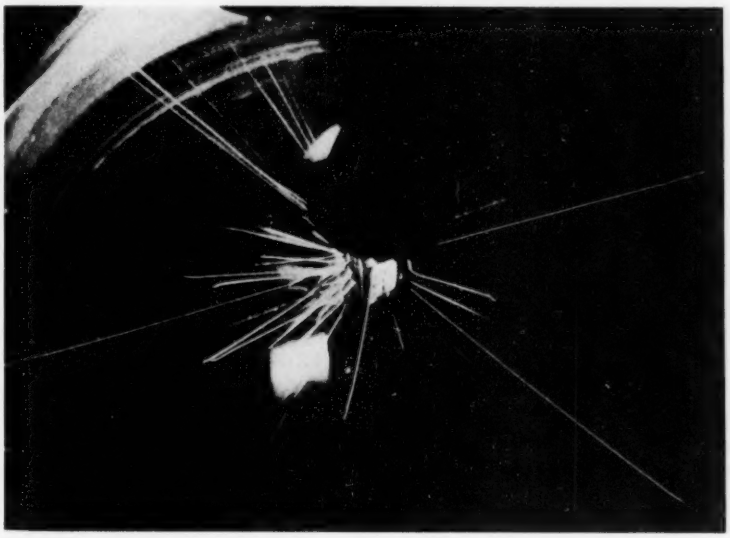


Fig. 15—Various tracks produced during bombardment of lithium by deutons. (Dee & Walton; *Proceedings*)

tentatively put Chadwick's value 1.0065 for it we get for $(T_1 - T_0)$ the value 16 millions of electron-volts. T_0 again is negligible, so that we are to compare half of this figure with the energy corresponding to the range 7.8 cm.—the right-hand end of the sloping part of the curve of Fig. 9—which is 8.3 millions. The agreement is entirely satisfactory. With boron the result is not so pleasing, for T_1 by equation (6) should be more than eleven millions, and two-thirds of this differs rather seriously from the energy-value corresponding to the end of the curve of Fig. 11, which is 6 millions. Kirchner got a photograph in which three coplanar tracks of the same appearance

diverge at mutual angles of 120° from a point in a boron target bombarded by protons, and Dee and Walton have noticed a number of trios of paths springing from such a target, but without being quite sure that they are not mere coincidences.[23]

Having now met with a case in which there may *not* be a balance between the two sides of such an equation as (6), we should now pause to inquire what can be done about such cases. Of course, such a disagreement might mean that the actual process is something entirely different from the one postulated in the equation, but it may not be necessary to make such a complete surrender of the theory. In equations (1) to (6), it is everywhere assumed that all the energy is retained by the material particles, in the form of kinetic energy or of rest-mass. Suppose that the process described by one of these equations, (6) for instance, is confirmed in every respect excepting that the final kinetic energy of the fragments is found to be less, by some amount Q, than the value of T_1 computed from the equation. One might then assume that the missing energy Q is radiated away in the form of one or more photons. Alternatively one might assume that the missing energy is retained by one of the material fragments in the form of "energy of excitation"; the rest-mass of the fragment, so long as it retained this energy and remained in the excited state, would then be correspondingly greater than its normal rest-mass, and the equation would be balanced if this abnormal value of mass were inserted into it in place of the normal one. Such explanations are frequently offered nowadays. They suffer, of course, from the disadvantage of being too easy; one can always postulate the necessary photons or excited states to explain any observed positive value of Q. But if they can ever be supported by independent proof of these excited states or photons, they will become much more convincing.

Lithium and boron are by far the best-studied of nuclei, in respect to their interactions with protons and deutons. It is true that our knowledge of the distribution-in-range curves of the fragments is still confined to comparatively low values of the energy of the bombarding particles, values less than 300,000 electron-volts. With higher energies it is to be presumed that the steps at the right-hand ends of the curves in Figs. 8 and 9 would move to the right, to the extent pre-

[23] If in the case of boron bombarded by protons it be assumed that two of the He nuclei fly off in directions making symmetrical angles $(\pi - \theta)$ and $(\pi + \theta)$ with the direction of the third, the distribution-in-θ of the disintegrations can be deduced from the curve of Fig. 14; it turns out that the most probable cases are those in which $\theta = 60°$ nearly, and all the three particles have nearly the same energy. A like deduction may be made for lithium bombarded by deutons, the neutron playing the part of third alpha-particle in the foregoing case; it is inferred that again the most probable types of disintegration are those in which all three share almost equally in the energy.

scribed by the increase of T_0 in equations (1) and (2); and so should the right-hand end of the sloping part of the curve in Fig. 9, and the extremity of the curve of Fig. 11. There is an indication of the first of these expected changes in the observation already quoted from Lewis Livingston and Lawrence, of 14.8-cm. fragments ejected from lithium by 1.33-MEV protons (page 133). We must wait for future data to test the others, and to see what happens to the heights of the steps and the general shape of the uninterpreted parts of the curves. Already however we have data bearing on the so-called "disintegration-function," or the relation of the total number of emitted fragments to the energy of the bombarding particles.

To speak of "total number of fragments" is to suggest too much. The present knowledge suffers from two limitations: the counts of fragments are made with apparatus which does not enclose the target completely and must be separated from the target by a screen, so that the fragments counted are only those which start off within a limited solid angle of deflections and have sufficient range to penetrate the screen. One generally makes a tentative correction for the former limitation, by assuming that the fragments go off equally in all directions and multiplying the number observed by the factor $4\pi/\omega$, where ω stands for the solid angle subtended by the detector as seen from the target. This factor may well be wrong, but perhaps does not vary seriously with the energy of the bombarding particles, so that at least the trend of the curve may not be distorted. For the latter limitation we have not the knowledge to make any allowance; it must always be stated that the count is of fragments having more than such-and-such a range, or such-and-such an energy. Every kind of device for observing transmutation suffers from some such lower limit, set either by the sensitivity of the device itself, or by the stopping-power of the wall which bounds it.

With their dense streams of protons and exceedingly thin films (page 113) Oliphant and Rutherford obtained the curves of Fig. 16: the disintegration-functions of lithium and boron, with respect to incident protons, up to proton-energies of some 200,000 electron-volts. The wall between the target and the gas of the ionization-chamber had an air-equivalent of 2.50 cm., and consequently the curves pertain only to fragments having ranges greater than this.[24] The rise from the axis is gradual, not abrupt; one might say that the shape of the curves suggests that the protons have, not a definite *capability* for transmuting which begins suddenly at a critical energy, but a *probability* of trans-

[24] I hear from Dr. Oliphant that the trend of the curve for the short-range fragments is just the same.

muting which increases smoothly from zero (though this suggestion might not occur to anyone not having foreknowledge of the current theory!). The least energy at which transmutation is observable should then depend entirely on the strength of the proton-stream and the sensitiveness of the apparatus; von Traubenberg, with a stream

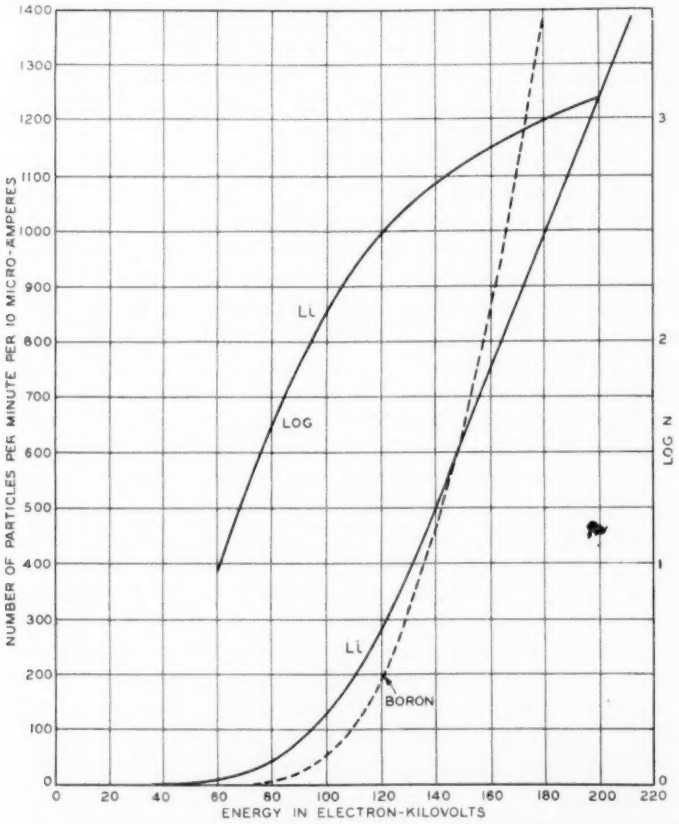


Fig. 16—Disintegration-functions of thin films of lithium and boron. (Oliphant & Rutherford)

perhaps as strong as that of Oliphant and Rutherford, observed one to three fragments per minute at 13,000 volts.

The curve of Fig. 17 extends very much further—all the way to 1.125 MEV—but was obtained with so thick a target of lithium (lithium fluoride, to be precise) that the protons came to a stop in the

mass, and the disintegrations observed at any voltage might have been produced by particles of any energy up to the maximum corresponding to the voltage. It comes from the Berkeley school, the data being procured chiefly by Henderson.[25] It refers only to fragments of ranges superior to 5.32 cm., a grave limitation, accepted in order to make sure that none of the primary protons could get into the detector (a Geiger counter). From 400,000 volts onward, the curve of Fig. 17 conforms

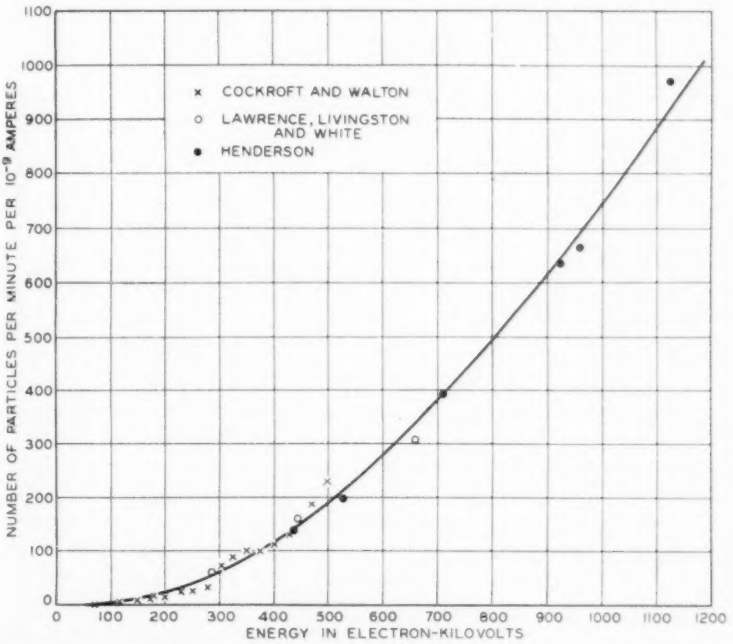


Fig. 17—Disintegration-function of lithium measured with a thick layer of lithium fluoride. (Henderson)

to a simple and somewhat surprising assumption: *viz.* the assumption that a proton of energy superior to 400,000 is neither more nor less efficient in disintegrating lithium than a proton of only 400,000 electron-volts, and that the whole of the rise in the curve from this voltage onwards is entirely due to the fact that the faster the proton, the farther it dives into the target and the more chances it has to

[25] The curve also fits the data of Cockcroft and Walton within the uncertainty of experiment, due regard being had to the difference in the values of the solid angle (letter from Dr. Henderson). In their work the screen between target and detector had an air-equivalent of 3 cm. (letter from Dr. Cockroft). The curve of Fig. 8 shows that this had the same effect as Henderson's 5.32 cm.

impinge on a nucleus before it is slowed down and its energy reduced beneath this particular value. The curve of Fig. 15 for lithium should then become horizontal at abscissa 400. At lower voltages, both curves concur in implying that the probability of disintegration depends on the energy of the proton. I will revert to this topic in a later article.

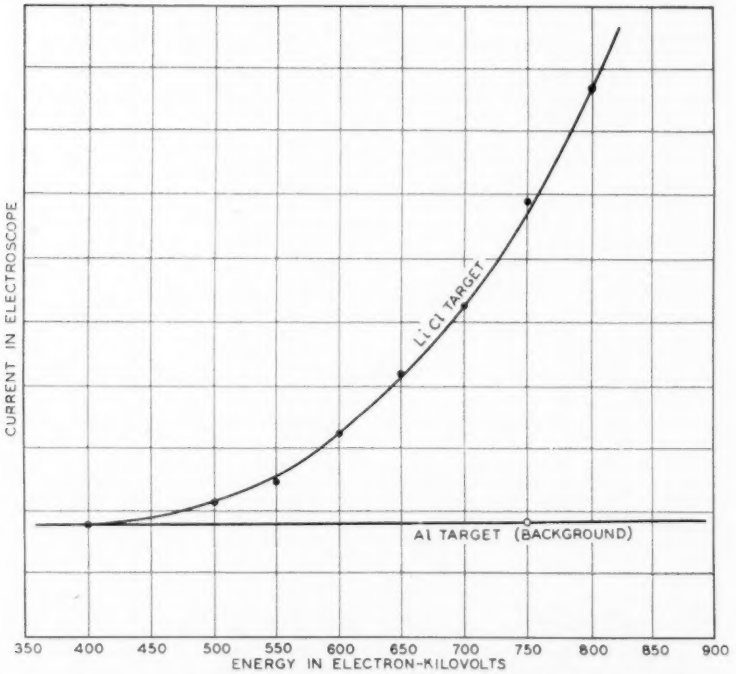


Fig. 18—Intensity of the mixture of neutrons and gamma-rays resulting when lithium is bombarded by deutons. (Crane & Lauritsen)

There are also modes of disintegration of lithium by deutons and by protons, in which neutrons and gamma-rays are emitted. These have been observed in Pasadena by Crane, Lauritsen and Soltan. Deutons are the more efficient of the two, but protons are sufficiently potent to have enabled Crane and Lauritsen to trace the curves of Fig. 18, in which the significant quantity is the difference between the ordinates of the two.[26] The ionization-chamber was walled inwardly

[26] Dr. Lauritsen writes me that the readings from which the lower curve is drawn were unchanged when the high voltage was removed; presumably therefore they represent the "background" due to the natural leaks of the electroscope. I am indebted to his letter for other as-yet-unpublished statements.

with paraffin, to accentuate the effect of the neutrons; it was however found that the readings were not considerably lessened when the paraffin coating was absent, and consequently Lauritsen infers that most of the effect is due to gamma-rays proceeding from the bombarded atoms. This inference is sustained by the fact that when the rays responsible for the effect are caused to pass through leaden screens, the ionization falls off exponentially with the thickness of the lead; and the value of the exponent suggests that the energy of the photons is about 1.5 MEV. One can easily think of a process whereby deutons might evoke neutrons from lithium nuclei:

$$ {}_1H^2 + {}_3Li^7 + T_0 = 2{}_2He^4 + {}_0n^1 + T_1, \qquad (7) $$

but with protons no plausible interaction comes readily to mind. Perhaps there is a two-stage process, the protons producing the reaction described by equation (1), the resultant He^4 nuclei striking other lithium nuclei and evoking neutrons. Or perhaps the neutrons and the gamma-rays alike result from the same processes as produce the groups of short-range alpha-particles revealed in Fig. 8. Questions of this intricate kind will probably predominate in the study of transmutation, in the years to come; and experiments on thin films will play a very important part in settling them, both because the likelihood of two-stage processes will be reduced, and because it may be possible to learn which isotopes are involved.

Little indeed is definitely known about the disintegration, by protons or deutons, of any other elements than lithium or boron. Charged fragments have been observed proceeding, in relatively small but yet appreciable number, from bombarded targets made of a great variety. But in many of these cases they may be due, so far as any of the observations tell, to a minute contamination of the target by boron derived from the glass of the enclosing tube; and the danger of this possible source of error was vividly brought out by Oliphant and Rutherford, when at first they observed such fragments, but ceased altogether to observe them when the original glass of their tube was replaced by a special boron-free variety! Beryllium and fluorine are the only elements, other than lithium and boron, of which these experimenters were sure of detecting fragments; for those of fluorine they were able to plot a disintegration-function and a distribution-in-range, which differed sufficiently in aspect from those of lithium and boron to exclude the possibility that these might be responsible; those of beryllium were too scanty for such tests. The elements with which they got no charged fragments, or only a few per minute, were the following: Fe, O, Na, Al, N, Au, Pb, Bi, Tl, U,

Th. But their observations were confined to protons of relatively low energy-values,—their upper limit was little over 200,000 electron-volts—and do not prove that faster particles are incapable of transmutation. The Berkeley school has already published a number of observations made with protons of energies ranging up to 710,000, and with deutons of energies attaining the unprecedented height of 3 MEV; and they find fragments in abundance from a wide diversity of targets.

Beryllium deserves a special paragraph, since it yields neutrons when bombarded, whether with alpha-particles from radioactive bodies; or with helium ions extracted from a discharge and endowed artificially with energies of 600,000 electron-volts and upward; or with deutons. The first of these processes is the one which led to the discovery of the neutron; the second, which incidentally marks the first employment of artificial alpha-particles (since these helium ions are alpha-particles in all but origin, except for the unimportant difference that each possesses an extra-nuclear electron while it is approaching the target) is a recent achievement of the Pasadena school (Crane, Lauritsen and Soltan); the third was achieved both at Pasadena and at Berkeley. These three processes are now in rivalry with one another, and it remains to be seen which will be producing the greatest number of neutrons, a year or five years hence. It is still very doubtful how the third takes place: perhaps the deuton merges with the beryllium nucleus, as in the other cases the alpha-particle is supposed to do (page 155), or perhaps it knocks a pre-existent neutron out of the beryllium structure and goes unaltered on its way. This too is a problem for the future, and one in the solving of which the charged fragments likewise observed will probably play a part.

The deuton itself is in all probability a complex particle; might it not be shattered in impinging against a nucleus, especially some heavy nucleus? This is the interpretation offered by Lawrence of the fact that in sending streams of deutons against targets of several different kinds, he observed charged fragments which were protons (not alpha-particles!) forming a group having a definite range and a definite energy not depending at all on the substance of the target. With 1.2-MEV deutons this characteristic energy of the protons is 3.6 MEV. A singular rule governs this quantity: if the energy of the bombarding particles is increased, that of the protons goes up by just the same amount—deutons of energy $(1.2 + x)$ MEV evoke protons of energy $(3.6 + x)$ MEV. The rule has been verified for values of x up to 1.8. Such a rule is just what one would expect, were there no other frag-

ments than the protons, excepting fragments of such great mass that they could take up the necessary momentum without taking an appreciable amount of kinetic energy. The heavy nucleus by itself is able to do this. However there are also neutrons, of which the energy is sufficient to let them be detected, and therefore by no means negligible. This is gratifying for the theory, inasmuch as if a proton is separated from a deuton, the residue should be a neutron (or else another proton and a free electron); but one is then obliged to assume that the neutron always takes the same kinetic energy, whatever that of the impinging deuton may have been. This seems rather odd, but nothing prohibits it. Streams of alpha-particles have been sent against compounds ("heavy water") containing deuterium in abundance, but as yet no neutrons have been detected coming off.

Transmutation by Impacts of Alpha-Particles [27]

Impact of an alpha-particle against a nucleus may result in the springing-off of one or more (or none) of four kinds of corpuscles: protons, photons, neutrons, positive electrons.

Transmutation with production of protons

This is the earliest-discovered type, of which I told at length in "Transmutation." The discovery was made by Rutherford in 1919 in experiments on nitrogen. At present the Cavendish school considers that this mode of transmutation has been proved for thirteen elements, none of atomic number greater than 19: the list comprises B, N, F, Ne, Na, Mg, Al, Si, P, S, Cl, A, K. The most frequently and fully studied cases are those of boron, nitrogen and aluminium.

The evidence that the fragments are protons is rather variegated. In some cases this has been proved by deflection-experiments; [28] recently it has been proved in some other cases by measuring both the range of the fragments and the ionization which they individually produce in a shallow chamber or a deep one (page 125); some observers are able to tell the scintillations due to protons from those which are due to alpha-particles.

Integral distribution-in-range curves of the fragments have been obtained for boron, nitrogen, fluorine, sodium, magnesium, aluminium and phosphorus. Most of them show more or less conspicuous plateaux, of which the most magnificent appear in the celebrated curves of Pose for aluminium, reproduced in "Transmutation"

[27] An expanded version of this section, with citations of additional data and reproductions of some curves, appears in the Physics Forum of the *Review of Scientific Instruments* for February 1934.

[28] "Transmutation," pp. 636–640, *B. S. T. J.*, Oct. 1931.

(Figs. 6, 7); from this there are all gradations of distinctness downward, ending with cases in which it is uncertain whether the ideal curve would be a smoothly-descending one, or would have a succession of short plateaux which in the actual curve are rounded off into indistinguishability.

By "ideal curve" in the foregoing sentence I mean, as heretofore (page 130), that which would be obtained with an infinitely narrow beam of fragments proceeding in a single direction and produced by alpha-particles all of a single speed and proceeding in a single direction. I must also add that many thousands of fragments should be counted, as otherwise the results are likely to be distorted by statistical fluctuations. It appears that in most of the experiments with bombarding alpha-particles, the departure from the ideal is much more considerable than in the best of the experiments with bombarding protons. The targets are usually so thick that the speeds of the alpha-particles vary considerably as they go through, and often so thick that these are swallowed up and every energy of bombarding particle, from the initial maximum down to zero, is represented among the impacts. This matters much more than it does with protons, because here the energy of the primary particles is often much greater than that of the fragments, and a small percentage variation of the former may entail a big one of the latter. The solid angles subtended by the exposed part of the target as seen from the source of the alpha-rays on the one hand, from the detector on the other, are frequently both large. This is particularly serious, because it appears that the ideal distribution-in-range curve would vary with the angle between the directions of the impinging particle and of the fragment. In some experiments the number of fragments observed has been too small to be immune to statistical fluctuations, and it is surprising that the plateaux in Pose's curves should be so clear despite this handicap.

Where two or more observers have studied a single element, there is generally enough concordance among their statements to assure the onlooker that at least the major groups of protons are recognizable. The prettiest case thus far is that of nitrogen: three researches on the integral distribution-in-range curve agree in showing a sharply-marked group of range about 17.5 cm (for protons ejected forward by full-speed alpha-particles from polonium, energy 5.3 MEV). The flattest plateau and sharpest step are to be seen in a curve by Chadwick Constable & Pollard, who approached very nearly to the ideal experiment in one respect, by using a stratum of nitrogen so thin that its air-equivalent was only 3 mm. All the protons of range superior to about 6 cm. belong to this group; there is another of inferior range, lately discovered

by Pollard. Phosphorus and sodium have been studied only by Chadwick Constable & Pollard, who find for the former a single group, for the latter a smoothly-descending integral curve which may betoken total absence of groups, or may be resolved, by some future and closer approach to the ideal curve, into a close succession of bends and corners. The four remaining elements—B, F, Mg, Al—show at least three groups apiece, and indeed Chadwick and Constable deduce four *pairs* of groups for aluminium and three for fluorine. To illustrate the degree of concurrence between different observers, I quote the values for the groups of aluminium—that is to say, values of the ranges of the protons belonging to these groups, ejected forward by 5.3-MEV alpha-particles —from the four authorities. Pose gives 28.5, 49.6, and 61.2 (cm of air-equivalent); Steudel, 33, 49, 63; M. de Broglie and Leprince-Ringuet, 30, 50, 60; Chadwick and Constable give 22, 26.5, 30.5, 34, 49, 55, 61, 66. More detailed comparisons had best be left to those who have practice in this field.

While nearly all of the data have been obtained by other methods than that of the expansion-chamber, a few beautiful pictures have been taken in which there appears the track of an alpha-particle passing through nitrogen, and this track is seen to end at a fork.[29] One of the tines of the fork is a long thin track, apparently that of a proton; there is only one other, and this is short and thick. It is inferred that these reveal the only fragments which there are, and that, in the usual though somewhat objectionable phrase, the alpha-particle has fused with the residual nucleus. The process is then expressed by the equation:

$$ {}_7\mathrm{N}^{14} + {}_2\mathrm{He}^4 + T_0 = {}_8\mathrm{O}^{17} + {}_1\mathrm{H}^1 + T_1, \tag{8} $$

the symbols being chosen according to the same principles as in equation (1). It is commonly assumed, though in no other case with such good evidence, that this happens in most if not in all cases, so that when a nucleus of atomic number Z and mass-number A is transmuted by an alpha-particle, the process often is:

$$ {}_Z\mathrm{M}^A + {}_2\mathrm{He}^4 + T_0 = {}_{Z+1}\mathrm{M}^{A+3} + {}_1\mathrm{H}^1 + T_1, \tag{9} $$

with an obvious symbolism. This is called "disintegration with capture" (though it is the case in which the objection to the name "disintegration," page 117, is gravest). The other conceivable case of "disintegration without capture" would be described thus:

$$ {}_Z\mathrm{M}^A + {}_2\mathrm{He}^4 + T_0 = {}_{Z-1}\mathrm{M}^{A-1} + {}_1\mathrm{H}^1 + {}_2\mathrm{He}^4 + T_1. \tag{10} $$

[29] "Transmutation," Figs. 10 and 11.

Disintegration-with-capture is very advantageous for the theorist, since when there are only two fragments after the interaction the principle of conservation of momentum suffices to determine the kinetic energy of either in terms of that of the other and that of the alpha-particle. In equation (9), T_0 stands for the kinetic energy of the alpha-particle, T_1 for the sum of the kinetic energies of the proton and the residual fragment, which call T_p and T_r respectively. Now excepting in the cloud-chamber experiments, it is only the proton which is detected, and therefore only T_p can be estimated from the data; but if the disintegration is by capture, then T_r and consequently T_1 can be deduced from T_0 and T_p. If however there are three or more final fragments, measurement of T_p is not sufficient to determine T_1. Also even in the case of disintegration-by-capture there will be uncertainty if the transmuted element is a mixture of two or more isotopes, since the value of T_r corresponding to an observed T_p will depend on the mass of the atom which is transmuted.

In a case of disintegration-by-capture, the simplest possible assumption is that $(T_1 - T_0)$ has a perfectly definite value, independent of T_0: there is conversion of a definite amount of kinetic energy into rest-mass (or vice versa), whatever the velocity of the alpha-particle may be. This may be tested by varying T_0; it may also be tested to some extent by observing protons ejected in various directions (relatively to the initial direction of the alpha-particles) since although the sum of T_p and T_r (which is T_1) should be the same for all of these protons those two quantities individually should vary, and T_p in particular should depend in a definite manner on the direction of the protons. Yet in nearly all such tests, the target is so thick that the alpha-particles impinging on various nuclei have very various speeds. How then shall we know which speed of proton to associate with which speed of alpha-particle, which value of T_p belongs with which of T_0? One naturally begins by assuming that the fastest of the primary particles produce the fastest of the protons. But plausible as this assumption seems at first, there are several cases known in which it is not true: cases in which a definite group of protons is evoked by alpha-particles of a definite interval of speeds, and neither faster nor slower particles are capable of producing them.

This phenomenon of "resonance," as it is called,[30] was first observed by Pose in the experiments on aluminium to which many pages were devoted in "Transmutation." It is evidently an important quality of nuclei, destined to be prominent in experiment and theory both.

[30] There is a tendency to use the term "resonance" to express the mere existence of groups, irrespective of whether they are evoked by alpha-particles of narrowly limited speeds. This is to be deprecated.

This makes it desirable to consider at some length how resonance may be detected. There are the following ways:

(*a*) When the target is thick, one may vary the energy K_0 which the particles possess when they strike the target-face K_0 (usually by varying the density of gas between the target-face and the source of the alpha-particles) and plot the integral distribution-in-range curve for many different values of K_0. Let us suppose that there is a certain proton-group evoked only by alpha-particles having energy between K_a and K_b, the notation being so chosen that $K_b < K_a < K_0$. Then it will be found that as K_0 is lowered, the step and plateau which reveal the group will remain unaltered until K_0 drops below a certain critical value (to be identified with K_a) after which they will fade out.

(*b*) In the foregoing conditions, one may use a very thin ionization-chamber and plot instead of the integral distribution-in-range curve a curve of the sort in Fig. 10, or the sort described on page 129 of which the ordinate stands for the number of fragments producing more than a certain chosen amount of ionization in the chamber. There will be various peaks in the curve corresponding to various groups, and if any of these is produced by "resonance" it will at first remain unaltered and then gradually disappear as K_0 is lowered.

(*c*) When targets thin enough to be completely traversed by the alpha-particles are available, one may leave K_0 unchanged and increase the thickness t of the target. The energies of the impinging particles in a target then vary from K_0 down to a minimum value K_1 which depends on t. If curves of any of the foregoing kinds be plotted for various values of K_1, and if any of the groups is produced by resonance, then the step or the peak corresponding to this group may be absent when K_1 is high (i.e. with the thinnest target) and will then make its appearance when K_1 is lowered past a certain critical value (again to be identified with K_a).

(*d*) If the target is so very thin that the loss of speed suffered by the alpha-particles in going through is negligible, and K_1 is sensibly equal to K_0, then when K_0 is varied the groups should appear and disappear when it becomes equal to K_a and K_b, respectively.

(*e*) Without subjecting the fragments to any analysis, one may simply measure the total number thereof (or rather, the total number having ranges superior to some fixed minimum) as function of K_0. Suppose the target to be thick; then, if all the proton-groups are evoked by resonance, the curve should display a sequence of steps and plateaux; if in addition to such there are groups which are evoked by particles of any energy over a wide interval, the steps need not vanish, but the plateaux should slope upward and may be curved.

If the target is very thin (in the sense of the previous paragraph) the curve ought to show a peak for each group. Such curves, by the usage of page 139, may be styled "disintegration functions" (the term "excitation-function" is also used).

(*f*) Finally, when the target is thick the mere existence of sharp steps in the integral distribution-in-range curves, may be taken as a suggestion of resonance, since if a group were evoked by alpha-particles of a wide range of energies it would probably have a broad distribution of speeds. But this is not a very strong argument by itself.

Despite this great variety of ways of testing for resonance, the situation is still confusing and confused.

Aluminium has been the object of most of the tests, doubtless because it figured in Pose's discovery. He used methods (*a*) and (*c*) and found resonance distinctly and even vividly displayed by the 60-cm. and the 50-cm. group, and not at all by the 25-cm. group. Chadwick and Constable used (*a*) and (*b*), and concluded that there is resonance for six at least of their eight groups, the two members of a pair appearing and disappearing together. (The remaining pair was elicited by alpha-particles of a limited interval of energy-values extending from a lower limit K_b to the highest value of K_0 which they had available.) They also used (*e*) with a very thin sheet of aluminium (air-equivalent 0.8 mm.) and got a curve with two well-defined peaks. But Steudel also had recourse to method (*e*), and the curve he got swept smoothly upward; it is true that his target was notably thicker (air-equivalent 5.2 mm.) and yet one would not expect such a thickness to blot out the peaks if they exist. Harder yet to explain away is the evidence of M. de Broglie and Leprince-Ringuet, who made test (*d*) with sheets of aluminium of air-equivalent 2.5 mm., and observed all three of Pose's groups over a wide range of values of K_0.—As for the other elements: boron and fluorine and magnesium have all been tested by method (*a*), and there are strong indications of resonance for all three, strongest for fluorine. Nitrogen has been studied by Pollard with a modification of (*e*), and he finds that resonance is displayed by the 6-cm. group but not by the stronger and better-known group of longer range.

Evidently this is a field which yearns for further cultivation, with more powerful sources of transmuting particles to make possible the use of narrower and more homogeneous beams of these, narrower pencils of fragments and thinner strata of matter. The discovery of the capacity of protons to transmute has probably diverted from it some of the attention which otherwise it would by now have received, but the lost ground will doubtless be made up in the course of years, after the

developments which that discovery has hastened shall have brought about the generation of streams of artificial alpha-particles more numerous by far than the natural ones. Meanwhile we must be content with scanty data and with fragmentary tests of the important question already mentioned: whether the energy transformed from rest-mass to vis viva or reversely—the quantity here denoted by $(T_1 - T_0)$, elsewhere commonly by Q, designated in German as the *Tönung* of the process—is a definite and characteristic quantity.

Certainly about resonance is essential to these tests; for if resonance exists, we have to correlate the energy of a group of protons with that particular energy of the alpha-particles which evokes the group; but if resonance does not occur, then probably the best we can do is to correlate the energy of the fastest of the ejected protons of a group with that of the fastest of the impinging particles—and if we make the latter guess when it ought not to be made, there will be trouble! Perhaps the most impressive evidence is that available for aluminium. Chadwick and Constable evaluated $(T_1 - T_0)$ for all of their eight groups: the six for which they demonstrated resonance, and the two which were evoked by alpha-particles of a limited interval of energies extending up to the highest which they used, which was 5.3 MEV. They find that $(T_1 - T_0)$ has a common value of $+2.3$ MEV for four of their groups—to wit, the longer-range members of their four pairs—and a common value of zero for the other four. Haxel plotted the integral distribution-in-range curves for the protons ejected by alpha-particles of several yet higher energies, running up almost to 9 MEV; he detected two groups; they did not display resonance, but he correlated the highest energy represented in each with the highest represented among the impinging particles, and he too found $+2.3$ MEV and zero for $(T_1 - T_0)$ in the two cases![31] Blackett analyzed eight examples of transmutation of nitrogen observed with the cloud-chamber (here he had the unique advantage of being able to observe the track of the residual nucleus and estimate its energy) and he reported for $(T_1 - T_0)$ a mean value of -1.27 MEV with a mean deviation of 0.42 from the mean. Future confirmation awaited this work also: Pollard, analyzing his integral distribution-in-range curves, made a computation of $(T_1 - T_0)$ for the 6-cm. group which exhibits resonance, and another for the 17.5-cm. group which does not, correlating in this latter case the energy of the fastest protons with that of the fastest alpha-particles; the results were -1.32 and -1.26 MEV.

[31] The precision of these values can hardly be estimated from what Chadwick and Constable say, but some idea of it can be gained from a graph in Haxel's article, *ZS. f. Phys.* **83**, p. 335 (1933), and *loc. cit.* footnote 27.

Such are the cases where there is the strongest proof for the twin doctrines that disintegration is by capture, and that a definite amount of energy is transformed between rest-mass and vis viva. The reader will have noticed in the latter case, that $(T_1 - T_0)$ appeared to be the same for a group which exhibits resonance and for another group which does not. This if certain may be taken to mean, that a particular group of protons—one may speak more graphically, and say: a particular proton in a particular level of the nitrogen nucleus—can be extracted by alpha-particles of a narrowly-limited range of energies between critical energy-values K_a and K_b, and can also be extracted by alpha-particles of *any* energy superior to a third critical value K_c which is greater than K_a and K_b. There is a good interpretation of this notion in the contemporary theory, which I reserve for the next article. It will also have been noticed that two different values of $(T_1 - T_0)$ were given for a single case, that of aluminium (there are also two for fluorine). This is to be taken as meaning that the residual nucleus may be left in either of two conditions, one of which may be the normal state, while the other must be an excited state (page 138). One then infers that the nucleus when left in the excited state will presently go over to the normal state, emitting a photon having an amount of energy equal to the difference between the two values of $(T_1 - T_0)$. It is very tempting to suppose that the gamma-rays known to be emitted from some elements during alpha-particle bombardments have this origin, but the measurements are not yet precise enough to prove this.[32]

In a case of disintegration-by-capture, the residual nucleus denoted by ${}_{z+1}\mathrm{M}^{A+3}$ in equation (9) might or might not be exactly the same as the nucleus of the known chemical atom (if such there be) of atomic number $(Z + 1)$ and mass-number $(A + 3)$. Can this be tested by comparing the rest-mass of the former with the mass of the latter as measured by Aston or Bainbridge? Unfortunately nothing of value can be concluded unless the atoms ${}_{z+1}\mathrm{M}^{A+3}$ and ${}_{z}\mathrm{M}^{A}$ have both had their masses determined with an accuracy permitting them to appear in the Table on page 109; and on inspecting this table one finds (with some surprise) that this is true for only one of the known processes, *viz.* the transmutation of fluorine. Assuming disintegration to be with capture, the process would be the following:

$$ {}_{9}\mathrm{F}^{19} + {}_{2}\mathrm{He}^{4} = {}_{10}\mathrm{Ne}^{22} + {}_{1}\mathrm{H}^{1} + (T_1 - T_0) \qquad (11) $$

Putting for $(T_1 - T_0)$ the value +1.67 MEV given by Chadwick and Constable, and for the rest-masses of the nuclei the values given in the

[32] Heidenreich has analyzed the data for boron, and concludes that they permit of this interpretation. (*ZS. f. Phys.* **87**, 675–693; 1933.)

Table, we get 23.002 for the left-hand member and 23.0043 for the right-hand member. The agreement is within the uncertainty of the data; so also would it have been, had $(T_1 - T_0)$ been ignored. Its importance is perhaps enhanced by the fact that it is *ex post facto:* the mass of Ne^{22} was inaccurately known at the time of the experiments of Chadwick and Constable, and there was ostensibly a disagreement.

I repeat that it is not proved that transmutation occurs in every case by capture; and an isolated value of $(T_1 - T_0)$, such as one often sees computed from a single observation on a particular group evoked by a particular beam of alpha-particles, is not necessarily valid.

Transmutation with production of neutrons

This mode of transmutation has been proved, according to the Cavendish school and the Joliots, for the elements Li, Be, B, F, Ne, Na, Mg, and Al. The outstanding cases are those of beryllium and boron, with lithium and fluorine following after. Negative results have been reported by the Joliots for H, C, O, N, P and Ca, and there is no record of a positive result for He. Positive results have been reported for quite a number of elements both light and heavy by the Vienna school.

There is nothing which can properly be called a distribution-in-range curve for neutrons; but there is something which is potentially as useful—the integral distribution-in-range curve of the protons emanating from a thin layer of matter rich in hydrogen, placed between the source of the neutrons and the detector. If one can measure the speed of a proton recoiling in a known direction from the impact of a neutron, one can deduce the speed of the neutron; in particular, if one can measure the speeds of the protons projected straight forward by central impacts of the oncoming neutrons, one may consider their speeds as practically the same as those of the neutrons themselves.[33] It is thus a proper procedure to obtain the integral distribution-in-range curve of the protons projected forward, and convert it into a distribution-in-energy curve which is that of the protons and the neutrons alike. It has however not been an easy procedure, on account of the sparseness of the available sources of neutrons and hence of the streams of recoiling protons. Chadwick has published[1] a solitary curve of this sort, relating to the neutrons from beryllium ejected by the alpha-particles of polonium; and Dunning has obtained a curve displaying good plateaux and steps, relating to the neutrons from beryllium ejected by yet faster alpha-particles.[34] Steps and plateaux, as heretofore, signify groups of protons and consequently groups of neutrons. Feather has

[33] Cf. Part I, page 300.

[34] To be published in the article mentioned in Footnote 27, and by Dr. Dunning himself.

achieved the feat of taking and examining no fewer than 6900 cloud-chamber photographs in order to deduce the distribution-in-speed of neutron-streams from the tracks of the recoiling nuclei of various kinds of atoms. Most observers publish no curves, but give only verbal accounts in which they state the thickness (in air-equivalent) of the intercepting screens athwart the proton-beam, for which they observed a notable falling-off of the strength of that beam; or else they state what groups they believe in, inferring them presumably from observations of that type. This makes tiresome and unsatisfactory reading.

Much of recent research is meant to detect the very fastest neutrons emitted from a given element, for a reason which will presently be obvious if it is not already. Chadwick gives 3.35 MEV for the energy of the fastest neutrons ejected from boron by polonium alpha-particles, and 12 MEV for those similarly ejected by beryllium, while Dunning gives 14.3 MEV for those which beryllium emits when bombarded by the somewhat faster alpha-particles from radon.

Curves called "disintegration-functions," or more commonly "excitation-functions," have been plotted several times for the neutrons from beryllium and once at least for those from boron. One must realize an important distinction between them and the curves obtained when the fragments are alpha-particles or protons, as in Figs. 16 and 17. When the fragments are charged particles, it is practically certain that *all* of them which reach the detector at all are duly detected. When the fragments are neutrons it is certain that the only ones detected are those which strike protons (or other nuclei) hard enough and squarely enough to give them a considerable amount of energy and enable them to produce a good many ions in the ionization-chamber; and it is equally certain that those constitute but a small fraction of the total number of neutrons, most of which go through the expansion-chamber unperceived. Would that this were at least a *constant* fraction! we could then rely on the shape of the so-called excitation-curve, while realizing that all its ordinates must be multiplied by some unknown but constant factor. But we must not suppose even this; it is practically certain that the factor varies with the speed of the neutrons, and hence in all probability with the speed of the primary alpha-particles; and hence the so-called excitation-curve must be distorted from the true curve of number-of-atoms transmuted *versus* energy-of-alpha-particles. (Also the distribution-in-range curves must be distorted.)

With these severe limitations in mind, one may consider the published excitation-functions. The most striking are those obtained with very thin films of beryllium, one by Chadwick and one by Bernardini,

which agree in showing a rather sudden rise of the curve from the horizontal axis, then a peak, then a valley and then a sweeping rise. It is hardly likely that the peak and the valley are entirely due to distortion of a truly smoothly-rising curve by the aforesaid agency; and the argument of paragraph (*e*) of page 149 leads us to infer a group of neutrons displaying resonance, in addition to other neutrons for which perhaps there is no resonance. Curves obtained with thick targets of beryllium or of boron have conspicuous steps, carrying the same implication. Those for boron (Chadwick and the Joliots) and some of those for beryllium (Rasetti, Bernardini) suggest but a single group, but there are other curves for beryllium suggesting two (in recent work of Chadwick's) and even four (Kirsch and Slonek). Thus, although the first four tests of resonance which I listed above (page 149) have as yet remained untried for emission of neutrons, the fifth has given some pretty convincing evidence in its favor.

It is always assumed that transmutation with emission of a neutron is a case of disintegration-by-capture, though no one has proof of this yet. The imagined process may be symbolized thus:

$$ {}_{Z}\mathrm{M}^{A} + {}_{2}\mathrm{He}^{4} + T_0 = {}_{Z+2}\mathrm{N}^{A+3} + {}_{0}n^{1} + T_1 \qquad (12) $$

Such equations as this are used for evaluating the rest-mass of the neutron, it being assumed that the rest-mass of the residual nucleus ${}_{Z+2}\mathrm{N}^{A+3}$ is identical with that of the nucleus of the atom of mass-number $(A + 3)$ and atomic number $(Z + 2)$. One encounters at once the difficulty that there are neutrons of a wide range of speeds, and consequently a wide range of values of T_1. It is necessary to assume that the slower neutrons leave behind them a nucleus in an excited state (page 138) and that only the very fastest leave behind them the normal nucleus which is to be identified with that of the isotope $(A + 3)$ of the element $(Z + 2)$. Doing this, Chadwick got consistent values for the mass of the neutron from the observations on boron and on lithium, assuming the nucleus M of equation (12) to be that of B^{11} and that of Li^{7} respectively.[35] To obtain a consistent value from the neutrons of beryllium, one would have to observe some at least having an energy as great as 12 MEV (when $T_0 = 5.3$ MEV). Those observed in the earlier work on beryllium were all much too slow. One of the driving motives of recent research has been the desire of finding at least a few of adequate energy; and it appears that this desire has at last been fulfilled.

[35] Were we to assume B^{10} and Li^{6}, the nucleus N would correspond to an isotope as yet unknown; this is a powerful but not an absolutely imperative argument against these choices. There is also the question of whether, if resonance occurs, the right correlation is being made between values of T_1 and values of T_0 (page 151).—The equation for the transmutation of boron has been worked out in Part I., pp. 323–324.

To guess at the total number of neutrons emitted (say) from beryllium it is necessary to know the excitation-curve and to make an estimate of the factor aforesaid. I confine myself to quoting from Chadwick: "The greatest effect is given by beryllium, where the yield is probably about 30 neutrons for every million alpha-particles of polonium which fall on a thick layer."

Transmutation with production of positive electrons

This mode of transmutation, as I mentioned earlier, has been observed by the Joliots with Be, B and Al, the primary corpuscles being polonium alpha-particles. Nothing has yet been published about distribution-in-range or disintegration-function. Positive electrons of energy as high as 3.1 MEV have been observed proceeding from aluminium.

Aluminium thus affords a case of an atom which under alpha-particle bombardment may emit from its nucleus a particle of any of three kinds: a proton, a neutron, a positive electron. It has been suggested by Joliot that there is actually only one process, in which a proton emerges either intact, or else split into a neutron and a positive electron which are its hypothetical components. If this can be verified it will have important bearings on various fundamental questions, including that of the mass of the neutron.[36] Boron also emits particles of all three kinds, but here the situation is complicated by the possibility that not all of the three proceed from the same isotope.

Transmutation by Neutrons

Transmutation by neutrons has been observed only with the Wilson chamber, and therefore rarely: there are a few scores of recorded cases, the fruit of twenty or thirty thousand separate photographs taken some by Feather at the Cavendish, some by Harkins and his colleagues at Chicago. What is observed is a pair of tracks diverging from a point in the midst of the gas contained in the chamber; it is inferred that the (invisible) path of a neutron extends from the neutron-source to the point of the divergence, and that the observed tracks are those of two fragments of a nucleus which that particle has struck. "Fragment" must be taken in the generalized sense of page 117: the substance of the neutron may be comprised in either or both of the two. Each case must be separately analyzed, taking into account the directions and the ranges of the fragments (it is here that the question of the range-*vs*-energy relations of massive nuclei, footnote 20, becomes crucial). It is possible to infer that in many cases the neutron is ab-

[36] See the reference in Footnote 27.

sorbed into the fragments—"disintegration with capture"—and even to estimate $(T_1 - T_0)$, which turns out to be usually if not always negative. There are some difficulties here, since in certain cases the process which is observed seems to be the converse of one of the well-known processes of generating neutrons, and yet $(T_1 - T_0)$ does not appear to have values equal in magnitude and opposite in sign for the two. The most startling feature of transmutation by neutrons is, that it occurs with nuclei which seem to be immune to other transmuting agents, notably carbon and oxygen. Other elements with which it occurs are nitrogen, fluorine, neon, chlorine and argon.

ACKNOWLEDGMENTS

I am greatly indebted to Monsieur F. Joliot, Professor E. O. Lawrence, Dr. J. R. Dunning and Dr. P. I. Dee for providing me with prints of several of the photographs which appear in this article (Figs. 1, 4, 5, 6, 14, 15); and to Dr. Dunning for criticism and advice in respect to several sections of the text.

REFERENCES

Transmutation by Protons and Deutons

Cavendish school:

J. Cockcroft & E. T. S. Walton: *Proc. Roy. Soc.* **A129**, 477–489 (1930); **136**, 619–630 (1932); **137**, 229–242 (1932).
P. I. Dee: *Nature* **132**, 818–819 (25 Nov. 1933).
P. I. Dee & E. T. S. Walton: *Proc. Roy. Soc.* **A141**, 733–742 (1933).
M. L. E. Oliphant & E. Rutherford: *Proc. Roy. Soc.* **A141**, 259–281 (1933). The same with R. B. Kinsey: *ibid.* 722–733.

Berkeley school:

M. C. Henderson: *Phys. Rev.* (2) **43**, 98–102 (1933).
E. O. Lawrence & M. S. Livingston: *Phys. Rev.* (2) **40**, 19–35 (1932).
Letters and abstracts by E. O. Lawrence, M. S. Livingston, M. G. White, G. N. Lewis, M. C. Henderson: *Phys. Rev.* (2) **42**, 150–151, 441–442 (1932); **43**, 212, 304–305, 369 (1933); **44**, 55–56, 56, 316–317, 317, 781–782, 782–783 (1933).

Other schools:

H. R. Crane, C. C. Lauritsen & A. Soltan, *Phys. Rev.* (2) **44**, 514 (1933) (effect of He+ ions); *ibid.* 692–693; Crane & Lauritsen, *ibid.* 783–784; **45**, 63–64 (1934).
C. Gerthsen: *Naturwiss.* **20**, 743–744 (1932).
F. Kirchner: *Phys. ZS.* **33**, 777 (1932); **34**, 777–786 (1933); with H. Neuert, **34**, 897–898 (1933). *Sitzungsber. d. kgl. Bayrischen Akad.* 129–134 (1933). *Naturwiss.* **21**, 473–478, 676 (1933).
H. Rausch v. Traubenberg, R. Gebauer, A. Eckart: *Naturwiss.* **21**, 26 (1933); *ibid.* 694.

Transmutation by Alpha-Particles

Transmutation with emission of protons:

P. M. S. Blackett: *Proc. Roy. Soc.* **A107**, 349–360 (1925).
W. Bothe: *ZS. f. Phys.* **63**, 381–395 (1930); *Atti del convegno di fisica nucleare*, Roma, 1932.
W. Bothe & H. Fränz: *ZS. f. Phys.* **43**, 456–465 (1927); **49**, 1–26 (1928).

W. Bothe & H. Klarmann: *Naturwiss.* **35**, 639–640 (1933).
M. de Broglie & L. Leprince-Ringuet: *C. R.* **193**, 132–133 (1931).
J. Chadwick, J. E. R. Constable & E. C. Pollard: *Proc. Roy. Soc.* **A130**, 463–489 (1931).
J. Chadwick & J. E. R. Constable: *Proc. Roy. Soc.* **A135**, 48–68 (1932).
K. Diebner & H. Pose: *ZS. f. Phys.* **75**, 753–762 (1932).
W. D. Harkins: with R. W. Ryan, *J. Am. Chem. Soc.* **45**, 2095–2107 (1923); with H. A. Shadduck, *Proc. Nat. Acad. Sci.* **2**, 707–714 (1926); with A. E. Schuh, *Phys. Rev.* (2) **35**, 809–813 (1930).
O. Haxel: *ZS. f. Phys.* **83**, 323–337 (1933).
F. Heidenreich: *ZS. f. Phys.* **86**, 675–693 (1933).
G. Hoffmann: *ZS. f. Phys.* **73**, 578–579 (1932).
C. Pawlowski: *C. R.* **191**, 658–660 (1930).
E. C. Pollard: *Proc. Roy. Soc.* **A141**, 375–385 (1933).
H. Pose: *Phys. ZS.* **30**, 780–782 (1929); **31**, 943–945 (1930). *ZS. f. Phys.* **60**, 156–167 (1930); **64**, 1–21 (1930); **67**, 194–206 (1931); **72**, 528–541 (1931). With F. Heidenreich: *Naturwiss.* **21**, 516–517 (1933).
E. Steudel: *ZS. f. Phys.* **77**, 139–156 (1932).
Additional early references given at the end of *Transmutation.*

Transmutation with emission of neutrons:

G. Bernardini: *ZS. f. Phys.* **85**, 555–558 (1933).
J. Chadwick: *Proc. Roy. Soc.* **A142**, 1–25 (1933).
N. Feather: *Proc. Roy. Soc.* **A142**, 689–714 (1933).
F. Joliot & I. Curie: *J. de Phys.* (7) **4**, 278–286 (1933).
G. Kirsch & W. Slonek: *Naturwiss.* **21**, 62 (1933).
F. Rasetti: *ZS. f. Phys.* **78**, 165–168 (1932).

Transmutation with emission of positive electrons:

I. Curie & F. Joliot: *J. de Phys.* (7) **4**, 494–500 (1933).

Transmutation by Neutrons

N. Feather: *Proc. Roy. Soc.* **A136**, 703–727 (1932); **142**, 689–709 (1933).
W. D. Harkins, D. M. Gans & H. W. Newson: *Phys. Rev.* (2) **44**, 529–537 (1933).
Letters and abstracts by W. D. Harkins, D. M. Gans, H. W. Newson: *Phys. Rev.* (2) **43**, 208, 362, 584, 1055 (1933); **44**, 236, 310, 945 (1933).
F. N. D. Kurie: *Phys. Rev.* (2) **43**, 771 (1933).

Abstracts of Technical Articles from Bell System Sources

Attenuation of Overland Radio Transmission in the Frequency Range 1.5 to 3.5 Megacycles per Second.[1] C. N. ANDERSON. Data on the effect of land upon radio transmission have been obtained during the past few years in connection with various site surveys. These data are for the general frequency range 1.5 to 3.5 megacycles per second and for various combinations of overwater and overland transmission as well as entirely overland. The generalizations in this paper are chiefly in the form of curves which enable one to make approximations of field strengths to be expected under the conditions noted above. The relation of these data to transmission in the broadcast frequency range is shown, and from the over-all picture, curves are developed which enable field strength estimates to be made for overland transmission in the extended frequency range.

The Radio Patrol System of the City of New York.[2] F. W. CUNNINGHAM and T. W. ROCHESTER. The application of radiotelephony to municipal police work in New York City is described from the organization viewpoint. Brief references are made to historical backgrounds and description of apparatus, and the steps taken to select a receiver suitable for local conditions are outlined. The method of controlling the patrol force by radio is described at some length with examples, and a summary of results during the first year is given to show the value of this means of communication to police work.

Electrical Disturbances Apparently of Extraterrestrial Origin.[3] KARL G. JANSKY. Electromagnetic waves of an unknown origin were detected during a series of experiments on atmospherics at high frequencies. Directional records have been taken of these waves for a period of over a year. The data obtained from these records show that the horizontal component of the direction of arrival changes approximately 360 degrees in about 24 hours in a manner that is accounted for by the daily rotation of the earth. Furthermore the time at which these waves are a maximum and the direction from which they come at that time changes gradually throughout the year in a way that is accounted for by the rotation of the earth about the

[1] *Proc. I. R. E.*, October, 1933.
[2] *Proc. I. R. E.*, September, 1933.
[3] *Proc. I. R. E.*, October, 1933.

sun. These facts lead to the conclusion that the direction of arrival of these waves is fixed in space; i.e., that the waves come from some source outside the solar system. Although the right ascension of this source can be determined from the data with considerable accuracy, the error not being greater than ± 7.5 degrees, the limitations of the apparatus and the errors that might be caused by the ionized layers of the earth's atmosphere and by attenuation of the waves in passing over the surface of the earth are such that the declination of the source can be determined only approximately. Thus the value obtained might be in error by as much as ± 30 degrees.

The data give for the coordinates of the region from which the waves seem to come a right ascension of 18 hours and a declination of − 10 degrees.

A Precision, High Power Metallographic Apparatus.[4] FRANCIS F. LUCAS. In 1927 the design of an advanced type of metallographic apparatus became of interest. Preliminary designs were prepared and discussed at a conference in Jena, Germany, with the scientific staff of Carl Zeiss. The Zeiss works was commissioned to construct the apparatus. The work was directed by Professor A. Kohler, an outstanding authority on the optics of the microscope, head of the mikro-department of the Zeiss works, and Professor Walter Bauersfeld, a director of the Zeiss Foundation and inventor of the Planetarium.

In this paper the author discusses the considerations which led to the design and describes the construction of the apparatus. It is the largest and the most powerful metallurgical microscope ever constructed. Capable of yielding crisp, brilliant images at magnifications of 4000 to 6000 diameters, the design required great mechanical stability, freedom from creep, absolute freedom from outside disturbances, the means to illuminate the specimen with light of any selected wave-length or group of wave-lengths within the visible spectrum and the highest order of achievement in optical equipment.

[4] Published in abridged form in *Metal Progress*, October, 1933.

Contributors to this Issue

H. S. Black, B.S. in Electrical Engineering, Worcester Polytechnic Institute, 1921. Western Electric Company, Engineering Department, 1921–25; Bell Telephone Laboratories, 1925–. Mr. Black's work has had to do with the development of carrier telephone systems.

Arthur G. Chapman, E.E., University of Minnesota, 1911. General Electric Company, 1911–13. American Telephone and Telegraph Company, Engineering Department, 1913–19, and Department of Development and Research, 1919–. Mr. Chapman is in charge of a group engaged in developing methods for reducing crosstalk between communication circuits, both open wire and cable, and evaluating effects of crosstalk on telephone and other services.

Karl K. Darrow, B.S., University of Chicago, 1911; University of Paris, 1911–12; University of Berlin, 1912; Ph.D., University of Chicago, 1917. Western Electric Company, 1917–25; Bell Telephone Laboratories, 1925–. Dr. Darrow has been engaged largely in writing on various fields of physics and the allied sciences.

Frederick B. Llewellyn, M.E., Stevens Institute of Technology, 1922; Ph.D., Columbia University, 1928. Western Electric Company, 1923–25; Bell Telephone Laboratories, 1925–. Dr. Llewellyn has been engaged in the investigation of special problems connected with radio and vacuum tubes.